STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 144403

TO: Minh-Tam Davis **Location: REM-3A24/3C18** **Art Unit: 1642** **Monday, February 23, 2004** **Case Serial Number: 10085108**	**From: Toby Port** **Location: Biotech-Chem Library** **Remsen 1A59** **Phone: 571-272-2523** **toby.port@uspto.gov**

Search Notes

Dear Examiner Davis,

Here are the results of your search.
Please feel free to contact me if you have any questions.

Toby Port

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 17:27:16 ; Search time 4182 Seconds
(without alignments)
10183.386 Million cell updates/sec

Title: US-10-085-108-21
Perfect score: 1041
Sequence: 1 ATGCCTCTCTTTCCAAACCT.....CTATCCAGTATCATCCCTAG 1041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1041	100.0	124497	9	HS232G24	AL022152 Human DNA
2	c	622.8	59.8	141672	9	HS142F18	AL031073 Human DNA
3		607	58.3	1932	9	AF490508	AF490508 Homo sapi
4		554	53.2	37757	9	HS326L12	AL023279 Human DNA
5	c	534.4	51.3	116210	9	HS406C18	AL023773 Human DNA
6		533	51.2	533	9	AF333706	AF333706 Homo sapi
7		486.4	46.7	1734	9	BC005891	BC005891 Homo sapi
8		486.4	46.7	1983	6	AR243321	AR243321 Sequence
9		486.4	46.7	1983	6	BD106883	BD106883 Isolated
10		486.4	46.7	1983	9	AF151378	AF151378 Homo sapi
11		486.4	46.7	1983	9	AF196482	AF196482 Homo sapi
12		486.4	46.7	2887	9	AF196483	AF196483 Homo sapi
13		486.4	46.7	2887	9	AF239802	AF239802 Homo sapi
14		486.4	46.7	2940	6	AR243322	AR243322 Sequence
15		486.4	46.7	2940	6	BD106884	BD106884 Isolated
16		486.4	46.7	3272	9	AF116195	AF116195 Homo sapi
17		484.8	46.6	1958	9	AF116194	AF116194 Homo sapi
18		484.8	46.6	2025	9	BC013318	BC013318 Homo sapi
19		411.4	39.5	4031	6	AR243307	AR243307 Sequence
20		411.4	39.5	4031	6	BD106868	BD106868 Isolated
21		411.4	39.5	4225	6	AR243314	AR243314 Sequence
22		411.4	39.5	4225	6	BD106875	BD106875 Isolated
23		411.4	39.5	4265	6	AR171862	AR171862 Sequence
24		411.4	39.5	4265	9	AF056334	AF056334 Homo sapi
25		411.4	39.5	6471	9	AF064589	AF064589 Homo sapi
26		409.8	39.4	4488	9	BC025969	BC025969 Homo sapi
27		399	38.3	274128	2	AC144277	AC144277 Macaca mu
28		342.2	32.9	1528	9	BC004105	BC004105 Homo sapi
29		342.2	32.9	2559	6	AR070825	AR070825 Sequence
30		342.2	32.9	2559	6	AR120096	AR120096 Sequence
31		342.2	32.9	2559	6	AR142176	AR142176 Sequence
32		342.2	32.9	2559	6	AR147416	AR147416 Sequence
33		342.2	32.9	2559	6	AR267625	AR267625 Sequence
34		342.2	32.9	2559	6	AX026470	AX026470 Sequence
35		342.2	32.9	2559	6	BD085552	BD085552 MAGE-10 e
36		342.2	32.9	3510	6	AR167380	AR167380 Sequence
37		342.2	32.9	3510	6	AX026469	AX026469 Sequence
38		342.2	32.9	3510	9	HSU10685	U10685 Human MAGE-
39	c	339	32.6	169351	9	AC116666	AC116666 Homo sapi
40		339	32.6	178515	2	AC009621	AC009621 Homo sapi
41		323.2	31.0	957	9	BT007340	BT007340 Homo sapi
42		323.2	31.0	957	12	BT008142	BT008142 Synthetic
43		323.2	31.0	1801	9	BC012744	BC012744 Homo sapi
44		323.2	31.0	1843	9	BC002455	BC002455 Homo sapi
45		321.6	30.9	72968	9	AC016941	AC016941 Homo sapi

ALIGNMENTS

RESULT 1
HS232G24
LOCUS HS232G24 124497 bp DNA linear PRI 09-MAR-2002
DEFINITION Human DNA sequence from clone RP6-232G24 on chromosome Xq27.1-27.3
Contains the gene for tha melanoma antigen gene family protein,
MAGEC3 and the MAGEC1 gene for melanoma antigen, family C,1,
complete sequence.
ACCESSION AL022152
VERSION AL022152.1 GI:3150086
KEYWORDS HTG; MAGEC1; MAGEC3.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 124497)

```
AUTHORS   Bird,C.
TITLE     Direct Submission
JOURNAL   Submitted (08-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
          Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
          humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On May 22, 1998 this sequence version replaced gi:2969932.
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations
          together with a note of the overlapping clone name. Note that the
          variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
          only a small overlap as described above.
          This sequence was finished as follows unless otherwise noted: all
          regions were either double-stranded or sequenced with an alternate
          chemistry or covered by high quality data (i.e., phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by at least
          one plasmid subclone or more than one M13 subclone; and the
          assembly was confirmed by restriction digest. The following
          abbreviations are used to associate primary accession numbers given
          in the feature table with their source databases: Em:, EMBL; Sw:,
          SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
          database can be found at
          http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
          was generated from part of bacterial clone contigs of human
          chromosome X, constructed by the Sanger Centre Chromosome X Mapping
          Group.  Further information can be found at
          http://www.sanger.ac.uk/HGP/ChrX
          RP6-232G24 is from the library RPCI-6 constructed by the group of
          Pieter de Jong. For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pPAC4
          This sequence is the entire insert of clone RP6-232G24 The true
          right end of clone RP3-326L12 is at 82755 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .124497
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP709G24232"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q27.1-27.3"
                     /clone="RP6-232G24"
                     /clone_lib="RPCI-6"
     repeat_region   6. .354
                     /note="L1MC2 repeat: matches 5195. .5546 of consensus"
     repeat_region   356. .762
                     /note="L1MC2 repeat: matches 5843. .6246 of consensus"
     repeat_region   763. .794
                     /note="16 copies 2 mer tg 93% conserved"
     repeat_region   946. .1161
                     /note="L1ME1 repeat: matches 5354. .5569 of consensus"
     repeat_region   1181. .1399
                     /note="L1ME1 repeat: matches 5914. .6136 of consensus"
     repeat_region   1471. .1606
                     /note="L2 repeat: matches 2574. .2708 of consensus"
     repeat_region   1939. .2330
                     /note="14 copies 28 mer 57% conserved"
     repeat_region   1969. .2326
                     /note="179 copies 2 mer gg 57% conserved"
     repeat_region   1972. .2331
                     /note="20 copies 18 mer 59% conserved"
     repeat_region   1996. .2265
                     /note="5 copies 54 mer 65% conserved"
     repeat_region   2029. .2181
                     /note="3 copies 51 mer 74% conserved"
     repeat_region   2153. .2347
                     /note="5 copies 39 mer 65% conserved"
     gene            2426. .4801
                     /gene="dA232G24.1"
     mRNA            join(2426. .2597,3291. .3384,3693. .4801)
                     /gene="dA232G24.1"
                     /product="dA232G24.1 (Melanoma antigen gene family
                     protein, MAGEC3)"
                     /note="(possible pseudogene)
                     match: cDNAs: Em:AF151378
                     match: ESTs: Em:BI460078"
                     /evidence=not_experimental
     repeat_region   2827. .2872
                     /note="23 copies 2 mer gg 78% conserved"
     CDS             3761. .4801
                     /gene="dA232G24.1"
                     /note="(possible pseudogene)
                     match: proteins: Sw:P43363 Tr:Q9BZ80 Tr:O60781 Tr:AAH04105
                     Tr:O95529 Tr:Q9R2A2 Sw:O15479 Tr:O89006 Tr:Q9BUN9
                     Tr:Q96M61"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="dA232G24.1 (Melanoma antigen gene family
                     protein, MAGEC3)"
                     /protein_id="CAA18146.1"
                     /db_xref="GI:3150087"
                     /translation="MPLFPNLPRLSFEEDFQNPSVTEDLVDAQDSIDEEEEDASSTSS
                     SSFHFLFPSSSSLSSSSPLSSPLPSTLILGVPEDEDMPAAGMPPLPQSPPEIPPQGPP
                     KISPQGPPQSPPQSPLDSCSSPLLWTRLDEESSSEEEDTATWHALPESESLPRYALDE
                     KVAELVQFLLLKYQTKEPVTKAEMLTTVIKKYKDYFPMIFGKAHEFIELIFGIALTDM
                     DPDNHSYFFEDTLDLTYEGSLIDDQGMPKNCLLILILSMIFIKGSCVPEEVIWEVLSA
                     IGVCAGREHFIYGDPRKLLTIHWVQRKYLEYREVPNSAPPRYEFLWGPRAHSEASKRS
                     LRVFIQAIQYHP"
     repeat_region   3876. .3959
                     /note="7 copies 12 mer 72% conserved"
     repeat_region   4027. .4110
                     /note="7 copies 12 mer 88% conserved"
     misc_feature    complement(5293. .5441)
                     /note="match: GSS: Em:AQ060261"
     repeat_region   6422. .6624
                     /note="MIR repeat: matches 47. .261 of consensus"
     repeat_region   8315. .8700
                     /note="MLT1B repeat: matches 1. .390 of consensus"
     repeat_region   9138. .9250
                     /note="MLT1B repeat: matches 1. .130 of consensus"
     repeat_region   9287. .9731
                     /note="L1MC3 repeat: matches 7166. .7618 of consensus"
     gene            11002. .16491
                     /gene="MAGEC1"
     mRNA            join(11002. .11086,11825. .11922,12078. .12184,12517. .16491)
                     /gene="MAGEC1"
                     /product="dA232G24.2 (Melanoma antigen gene family
                     protein, MAGEC1)"
                     /note="match: cDNAs: Em:AF056334
                     match: ESTs: Em:AI652057 Em:AL121366 Em:AI126114
                     Em:BG480822"
                     /evidence=not_experimental
     misc_feature    complement(11962. .12294)
                     /note="match: GSS: Em:AQ064665"
     CDS             join(12181. .12184,12517. .15941)
                     /gene="MAGEC1"
                     /note="match: proteins: Sw:O60732 Tr:O95529 Sw:Q9UBF1
                     Tr:Q96D45 Tr:Q99NC2 Tr:Q99PH7 Sw:Q9Y5V3"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="dA232G24.2 (Melanoma antigen gene family
                     protein, MAGEC1)"
                     /protein_id="CAD27434.1"
                     /db_xref="GI:20095260"
                     /translation="MGDKDMPTAGMPSLLQSSSESPQSCPEGEDSQSPLQIPQSSPES
                     DDTLYPLQSPQSRSEGEDSSDPLQRPPEGKDSQSPLQIPQSSPEGDDTQSPLQNSQSS
                     PEGKDSLSPLEISQSPPEGEDVQSPLQNPASSFFSSALLSIFQSSPESTQSPFEGFPQ
                     SVLQIPVSAASSSTLVSIFQSSPESTQSPFEGFPQSPLQIPVSRSFSSTLLSIFQSSP
                     ERTQSTFEGFAQSPLQIPVSPSSSSTLLSLFQSFSERTQSTFEGFAQSSLQIPVSPSF
                     SSTLVSLFQSSPERTQSTFEGFPQSPLQIPVSSSSSSTLLSLFQSSPERTHSTFEGFP
                     QSLLQIPMTSSFSSTLLSIFQSSPESAQSTFEGFPQSPLQIPGSPSFSSTLLSLFQSS
                     PERTHSTFEGFPQSPLQIPMTSSFSSTLLSILQSSPESAQSAFEGFPQSPLQIPVSSS
                     FSYTLLSLFQSSPERTHSTFEGFPQSPLQIPVSSSSSSSTLLSLFQSSPECTQSTFEG
                     FPQSPLQIPQSPPEGENTHSPLQIVPSLPEWEDSLSPHYFPQSPPQGEDSLSPHYFPQ
                     SPPQGEDSLSPHYFPQSPQGEDSLSPHYFPQSPPQGEDSMSPLYFPQSPLQGEEFQSS
```

```
                     LQSPVSICSSSTPSSLPQSFPESSQSPPEGPVQSPLHSPQSPPEGMHSQSPLQSPESA
                     PEGEDSLSPLQIPQSPLEGEDSLSSLHFPQSPPEWEDSLSPLHFPQFPPQGEDFQSSL
                     QSPVSICSSSTSLSLPQSFPESPQSPPEGPAQSPLQRPVSSFFSYTLASLLQSSHESP
                     QSPPEGPAQSPLQSPVSSFPSSTSSSLSQSSPVSSFPSSTSSSLSKSSPESPLQSPVI
                     SFSSSTSLSPFSEESSSPVDEYTSSSDTLLESDSLTDSESLIESEPLFTYTLDEKVDE
                     LARFLLLKYQVKQPITKAEMLTNVISRYTGYFPVIFRKAREFIEILFGISLREVDPDD
                     SYVFVNTLDLTSEGCLSDEQGMSQNRLLILILSIIFIKGTYASEEVIWDVLSGIGVRA
                     GREHFAFGEPRELLTKVWVQEHYLEYREVPNSSPPRYEFLWGPRAHSEVIKRKVVEFL
                     AMLKNTVPITFPSSYKDALKDVEERAQAIIDTTDDSTATESASSSVMSPSFSSE"
     repeat_region   12417. .12500
                     /note="7 copies 12 mer 70% conserved"
     repeat_region   12420. .12497
                     /note="13 copies 6 mer tcctcc 67% conserved"
     repeat_region   12675. .12752
                     /note="2 copies 39 mer 89% conserved"
     repeat_region   12721. .12912
                     /note="4 copies 48 mer 88% conserved"
     repeat_region   12928. .13977
                     /note="10 copies 105 mer 94% conserved"
     repeat_region   13844. .14719
                     /note="2 copies 438 mer 82% conserved"
     repeat_region   14023. .14262
                     /note="5 copies 48 mer 84% conserved"
     repeat_region   14039. .14266
                     /note="19 copies 12 mer 58% conserved"
     repeat_region   14217. .14306
                     /note="2 copies 45 mer 100% conserved"
     repeat_region   14455. .14742
                     /note="24 copies 12 mer 56% conserved"
     repeat_region   14480. .14719
                     /note="5 copies 48 mer 89% conserved"
     repeat_region   14683. .14997
                     /note="3 copies 105 mer 86% conserved"
     repeat_region   14948. .15049
                     /note="2 copies 51 mer 99% conserved"
     repeat_region   15169. .15222
                     /note="3 copies 18 mer 88% conserved"
     repeat_region   17466. .17718
                     /note="MIR repeat: matches 11. .261 of consensus"
     repeat_region   19287. .19670
                     /note="MLT1B repeat: matches 1. .390 of consensus"

  Query Match             100.0%;  Score 1041;  DB 9;  Length 124497;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-297;
  Matches 1041;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 3820

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 3880

Qy        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3881 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 3940

Qy        181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3941 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 4000

Qy        241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4001 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 4060

Qy        301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4061 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 4120

Qy        361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAGAGGAGGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4121 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAGAGGAGGAT 4180

Qy        421 ACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4181 ACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAA 4240

Qy        481 AAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4241 AAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACA 4300

Qy        541 AAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4301 AAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTC 4360

Qy        601 GGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4361 GGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCC 4420

Qy        661 GACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4421 GACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATT 4480

Qy        721 GATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4481 GATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATA 4540

Qy        781 AAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4541 AAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGT 4600

Qy        841 GCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4601 GCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTG 4660

Qy        901 CAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAATTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4661 CAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAATTT 4720

Qy        961 TTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAGAAGTCTTAGAGTTTTTATCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4721 TTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAGAAGTCTTAGAGTTTTTATCCAA 4780

Qy       1021 GCTATCCAGTATCATCCCTAG 1041
              |||||||||||||||||||||
Db       4781 GCTATCCAGTATCATCCCTAG 4801

RESULT 2
HS142F18/c
LOCUS       HS142F18              141672 bp    DNA     linear   PRI 22-NOV-2001
DEFINITION  Human DNA sequence from clone RP1-142F18 on chromosome Xq26.3-27.2,
            complete sequence.
ACCESSION   AL031073
VERSION     AL031073.1  GI:4090213
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 141672)
  AUTHORS   Pavitt,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 31, 1998 this sequence version replaced gi:3646045.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
```

chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/ChrX
RP1-142F18 is from the library RPCI-1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone RP1-142F18 It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true right end of clone RP1-142F18 is at 141672 in this sequence. The true left end of clone GS1-54N10 is at 139131 in this sequence. The true right end of clone RP3-406C18 is at 100 in this sequence.

FEATURES Location/Qualifiers

source	1. .141672
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="RZPD:RPCIP704F18142"
	/db_xref="taxon:9606"
	/chromosome="X"
	/map="q26.3-27.2"
	/clone="RP1-142F18"
	/clone_lib="RPCI-1"
misc_feature	complement(3934. .4358)
	/note="match: GSS: Em:AQ729073"
repeat_region	4956. .5027
	/note="6 copies 12 mer 75% conserved"
repeat_region	4958. .5026
	/note="23 copies 3 mer gag 75% conserved"
repeat_region	6704. .6727
	/note="12 copies 2 mer ca 100% conserved"
misc_feature	complement(7824. .7943)
	/note="match: STS: Em:L24631"
repeat_region	8429. .8500
	/note="36 copies 2 mer at 68% conserved"
repeat_region	8433. .8504
	/note="6 copies 12 mer 72% conserved"
misc_feature	12557. .12989
	/note="match: GSS: Em:AQ017378"
repeat_region	13342. .13401
	/note="5 copies 12 mer 80% conserved"
repeat_region	17074. .17317
	/note="122 copies 2 mer cc 55% conserved"
misc_feature	complement(19214. .19587)
	/note="match: GSS: Em:AQ596941"
repeat_region	19991. .20802
	/note="LTR1 repeat: matches 1. .785 of consensus"
repeat_region	21443. .21641
	/note="L1MA3 repeat: matches 6111. .6304 of consensus"
repeat_region	26327. .26581
	/note="L1PB3 repeat: matches 5894. .6150 of consensus"
repeat_region	27834. .28342
	/note="L1MA4A repeat: matches 5663. .6179 of consensus"
repeat_region	28637. .29340
	/note="L1M3 repeat: matches 4824. .5538 of consensus"
repeat_region	29537. .29572
	/note="3 copies 12 mer 100% conserved"
repeat_region	30190. .30271
	/note="41 copies 2 mer at 67% conserved"
repeat_region	30200. .30271
	/note="6 copies 12 mer 77% conserved"
repeat_region	33068. .33151
	/note="42 copies 2 mer ta 77% conserved"
repeat_region	33076. .33147
	/note="6 copies 12 mer 79% conserved"
repeat_region	33554. .33661
	/note="9 copies 12 mer 75% conserved"
repeat_region	33560. .33661
	/note="51 copies 2 mer at 73% conserved"
misc_feature	complement(35997. .36405)
	/note="match: GSS: Em:AQ354465"
misc_feature	complement(35998. .36419)
	/note="match: GSS: Em:AQ120433"
misc_feature	complement(35998. .36405)
	/note="match: GSS: Em:AQ353948"
misc_feature	complement(36099. .36409)
	/note="match: GSS: Em:AQ486967"
misc_feature	complement(36284. .36418)
	/note="match: GSS: Em:AQ002154"
misc_feature	complement(36289. .36417)
	/note="match: GSS: Em:AQ431535"
repeat_region	37079. .37111
	/note="LTR7 repeat: matches 1. .33 of consensus"
misc_feature	37111. .37242
	/note="match: GSS: Em:AQ263270"
misc_feature	37111. .37195
	/note="match: GSS: Em:AQ077230"
repeat_region	37195. .39609
	/note="HERVH repeat: matches 460. .2987 of consensus"
repeat_region	39604. .40459
	/note="HERVH repeat: matches 3140. .3993 of consensus"
repeat_region	40513. .41679
	/note="HERVH repeat: matches 3994. .5149 of consensus"
repeat_region	41681. .41900
	/note="HERVH repeat: matches 5600. .5819 of consensus"
repeat_region	41901. .42278
	/note="HERVH repeat: matches 7324. .7713 of consensus"
repeat_region	42279. .42710
	/note="LTR7 repeat: matches 1. .450 of consensus"
repeat_region	45952. .46111
	/note="THE1B-INTERNAL repeat: matches 1. .166 of consensus"
repeat_region	46102. .46169
	/note="THE1B-INTERNAL repeat: matches 97. .170 of consensus"
repeat_region	46147. .47366
	/note="THE1B-INTERNAL repeat: matches 310. .1574 of consensus"
repeat_region	47730. .48123
	/note="MSTC repeat: matches 1. .364 of consensus"
repeat_region	58971. .61286
	/note="L1PA15-16 repeat: matches -707. .1049 of consensus"
repeat_region	70284. .70609
	/note="L1MB1 repeat: matches 5828. .6162 of consensus"
repeat_region	70618. .70803
	/note="LTR2 repeat: matches 151. .338 of consensus"
repeat_region	72998. .73111
	/note="57 copies 2 mer at 78% conserved"
repeat_region	73019. .73114
	/note="8 copies 12 mer 80% conserved"
misc_feature	73925. .74189
	/note="match: GSS: Em:B34956"
repeat_region	74381. .74784
	/note="TIGGER1 repeat: matches 2014. .2418 of consensus"
misc_feature	74867. .74982
	/note="match: GSS: Em:AQ551643"
misc_feature	74919. .75076
	/note="match: GSS: Em:B80788"
repeat_region	74947. .75282
	/note="LTR2 repeat: matches 112. .449 of consensus"
repeat_region	75283. .83129
	/note="HERVE repeat: matches 1. .7812 of consensus"
misc_feature	83154. .83255
	/note="match: GSS: Em:AQ081161"
misc_feature	complement(83246. .83501)

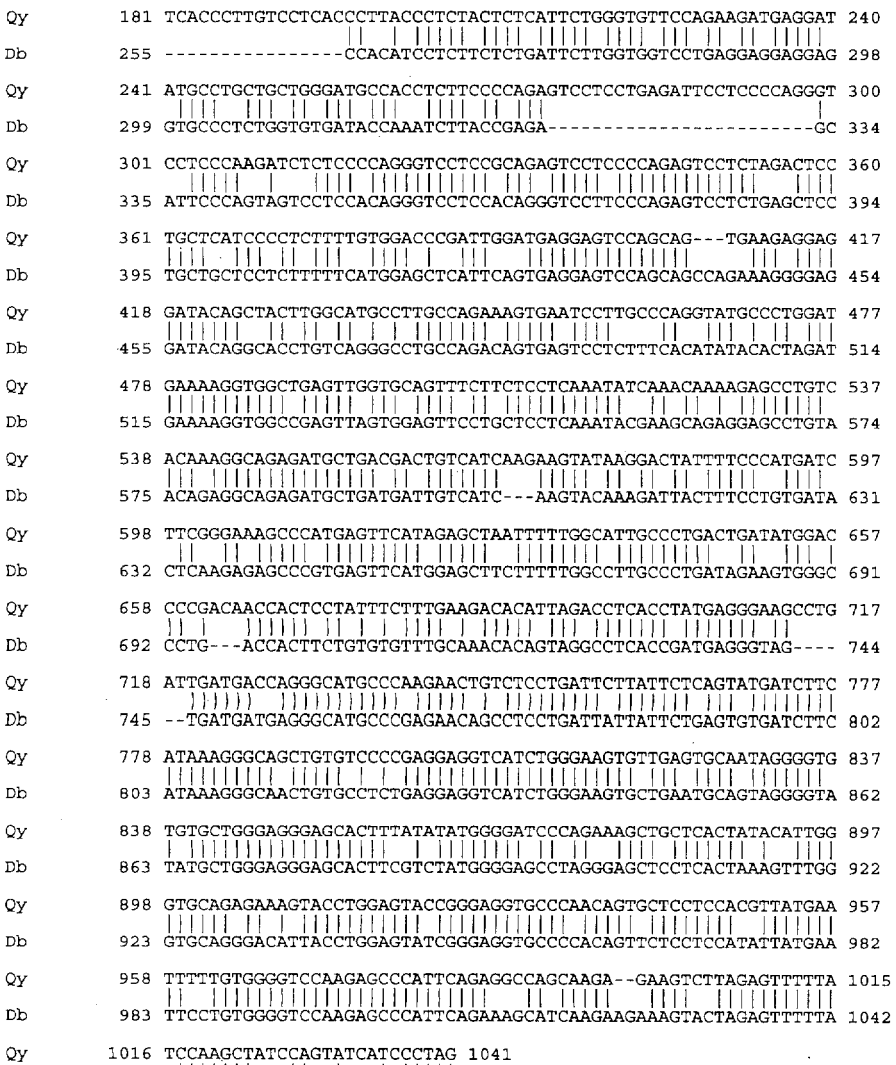
```
                /note="match: GSS: Em:AQ667065"
     repeat_region   83298. .83633
                /note="LTR2 repeat: matches 112. .449 of consensus"
     repeat_region   83634. .85637
                /note="TIGGER1 repeat: matches 1. .1994 of consensus"
     repeat_region   89169. .89216
                /note="24 copies 2 mer gt 93% conserved"
     misc_feature    complement(89444. .89909)
                /note="match: GSS: Em:B35662"
     repeat_region   95155. .95182
                /note="14 copies 2 mer tt 100% conserved"
     repeat_region   95228. .96236
                /note="L1PA3 repeat: matches 5138. .6146 of consensus"
     repeat_region   96760. .96951
                /note="16 copies 12 mer 80% conserved"
     repeat_region   99609. .99644
                /note="3 copies 12 mer 94% conserved"
     repeat_region   99611. .99642
                /note="16 copies 2 mer tc 96% conserved"
     repeat_region   101422. .101461
                /note="20 copies 2 mer tc 85% conserved"
     repeat_region   103090. .103133
                /note="22 copies 2 mer gg 75% conserved"
     misc_feature    103799. .103936
                /note="match: GSS: Em:AQ042099"
     repeat_region   104831. .104870
                /note="20 copies 2 mer tg 87% conserved"
     misc_feature    106895. .107214
                /note="match: GSS: Em:B70993"

  Query Match             59.8%;  Score 622.8;  DB 9;  Length 141672;
  Best Local Similarity   79.7%;  Pred. No. 4.5e-173;
  Matches  834;  Conservative    0;  Mismatches  152;  Indels   60;  Gaps    6;

Qy          4 CCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGTGTG 63
              |||||||| |||| |||||| ||||||| ||||||| ||||||||||| | | | ||
Db       5142 CCTCTCTTCCCAATCCTTCCCCGCCTCAACTTTGAGCAAGACTTCCAGACCGCAATTCTG 5083

Qy         64 ACAGAGGACTTGGTAGA---TGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              | | ||| |||||||||    |||||||||||||| ||| |||||||||||||| ||||||
Db       5082 ATACAGGTCTTGGTAGAATCTGCACAGGATTCCACAGAAGAGGAGGAGGAGGAAGCCTCC 5023

Qy        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
              |||| |||||||||||||||||||||||||||||||||||| ||||||||  |||||
Db       5022 TCCATTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTGCTCTTCCTCCTCC---- 4967

Qy        181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
                  |||  ||| |  |||||||||| |||||||| || |||||| ||  |||| ||||
Db       4966 ----------CTCTTCCTCATTCTCTACTCTGATTCTGGATGCTCCAGAGGACAAGGAC 4918

Qy        241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300
              |||||||||||||||| | |||  |||||||||||||| ||||   ||||||||  || |
Db       4917 ATGCCTGCTGCTGGGTTTCCAATTCTTCCCCAGAGTGCTCCCAGGATTCCTCTTCATG-- 4860

Qy        301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360
                                                  |||| |||||||  |||||   ||||
Db       4859 ---------------------------------ATCCTACCCAGAGACCTCTGAGCTCC 4834

Qy        361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAG 417
              |||||| | | |   ||||||   || |||||||||||||||||||     ||| ||||
Db       4833 TGCTCTTTCTCCTCTTTGTGAGGCCCATTGGATGAGGAGTCCAGCAGGCAAAAAGTGGAG 4774

Qy        418 GATACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGAT 477
              ||| ||||||| |||||||||||| |||| ||||||||||||| ||  |||||  ||||||
Db       4773 GATGCAGCTACATGGCATGCCTTGTCAGACAGTGAATCCTTGCTCACATATGCACTGGAT 4714

Qy        478 GAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTC 537
              |||||||| |||||||||||||||||||||||||||| ||||||| ||| |||||||||
Db       4713 GAAAAGGTGCCTGAGTTGGTGCAGTTTCTTCTCCTCAATTATCAAATAAAGGAGCCTGTC 4654

Qy        538 ACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATC 597
              |||||||| |||||||||||||||||||||||||||||||||||||||  |||||  |||||
Db       4653 ACAAAGGCTGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTGCTTTCCTGTGATC 4594

Qy        598 TTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGAC 657
              ||| ||||||| |||||||||| || ||||| |||||||||||||||||||| ||||||
Db       4593 TTCAGGAAAGCTCATGAGTTCATGGAACTAATGTTTGGCATTGCCCTGACTGAGATGGAC 4534

Qy        658 CCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTG 717
              || |||||||||||||| |||||||| ||||| |||| ||| | | | || || |||
Db       4533 ACCAACAACCACTCCTATGTCTTTGAAAACACAGTAGAACTCTGCGACCAAGGGAGGCTG 4474

Qy        718 ATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTC 777
              | ||||| |||||| ||| ||||||||    ||||| |||||||||||| ||| || |||||
Db       4473 AGTGATGGCCAGGGAATGTCCAAGAACCACCTCCTAATTCTTATTCTGAGTGTGGTCTTC 4414

Qy        778 ATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTG 837
              |||||||||||||||| | ||| ||||||| ||||||||| ||| | | |||||||||
Db       4413 ATAAAGGGCAGCTGTGCCTCTGAGAAGGTCATTTGGGAAGTGCTGAATTCGATAGGGGTG 4354

Qy        838 TGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGG 897
              | |||||||||||||| ||| || |||||| | |||||  ||| |||||||||| |||||
Db       4353 TATGCTGGGAGGGAGCCCTTCATTTATGGGAAGCCCAGGGAGCTCCTCACTATAGATTGG 4294

Qy        898 GTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAA 957
              |||||||||||||| |||||||||||||||||||| ||| ||||| ||||   ||||||
Db       4293 GTGCAGAGAAAGTATCTGGAGTACCGGGAGGTGCCCTACAATGCTCGTCCAGAGTATGAA 4234

Qy        958 TTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGA--GAAGTCTTAGAGTTTTTA 1015
              || ||||| |||| |||||||||||||||   |||||||   ||||| ||||| ||||
Db       4233 TTATTGTGAGGTCTAAGAGCCCATTCAGAAATCAGCAAGACGAAAGTCCTAGAGTGTTTA 4174

Qy       1016 TCCAAGCTATCCAGTATCATCCCTAG 1041
                | |||||  ||||| |||  ||||
Db       4173 GGC-AGCTAGACAGTACCATTGCTAG 4149


RESULT 3
AF490508
LOCUS       AF490508                1932 bp    mRNA    linear   PRI 04-APR-2002
DEFINITION  Homo sapiens hepatocellular carcinoma-associated protein HCA2 mRNA,
            complete cds.
ACCESSION   AF490508
VERSION     AF490508.1  GI:19919741
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1932)
  AUTHORS   Dong,X. and Chen,W.
  TITLE     Identification of genes in the chromosome X that are differentially
            expressed in hepatocellular carcinoma
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1932)
  AUTHORS   Dong,X. and Chen,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2002) Department of Immunology, T Cell Research
            Lab., Xueuan Road 38, Beijing 100083, China
FEATURES             Location/Qualifiers
     source          1. .1932
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
     CDS             1. .1932
                     /codon_start=1
                     /product="hepatocellular carcinoma-associated protein
                     HCA2"
                     /protein_id="AAM08355.1"
                     /db_xref="GI:19919742"
                     /translation="MLLPCHWVLDATFSDGSLGQWVKNTCATYALSPVVLPPQPQPRK
                     KATDKDYSAFHLGHLREVRLFLRGGTSDQRMDSLVLCPTYFKLWRTLSGSPGLQLSDL
                     HFGSQPEGKFSLRRAVSVKQREEPQDWPLNEKRTLWKDSDLPTWRRGTGYTLSLPAVS
                     PGKRLWGEKAGSLPESEPLFTYTLDEKVDKLVQFLLLKYQAKEPLTRAEMQMNVINTY
                     TGYFPMIFRKAREFIEILFGISLTEVDPDHFYVFVNTLDLTCEGSLSDEQGMPQNRLL
```

```
                     ILILSVIFIKGNCASEEVIWEVLNAIGPWSALAGFADVLSRLALWESEGPEAFCEESG
                     LRSAEGSVLDLANPQGLAGHRQEDGRRGLTEASPQQKKGGEDEDMPAAGMPPLPQSPP
                     EIPPQGPPKISPQGPPQSPPQSPLDSCSSPLLWTRLDEESSSEEEDTATWHALPESES
                     LPRYALDEKVAELVQFLLLKYQTKEPVTKAEMLTTVIKKYKDYFPMIFGKAHEFIELI
                     FGIALTDMDPDNHSYFFEDTLDLTYEGSLIDDQGMPKNCLLILILSMIFIKGSCVPEE
                     VIWEVLSAIGPIQRPAREVLEFLSKLSSIIPSAFPSWYMDALKDMEDRAQAIIDTTDD
                     ATAMASASPSVMSTNFCPE"
BASE COUNT      477 a    489 c    510 g    456 t
ORIGIN

  Query Match             58.3%;  Score 607;  DB 9;  Length 1932;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-168;
  Matches  607;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        228 AGAAGATGAGGATATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGAT 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1122 AGAAGATGAGGATATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGAT 1181

Qy        288 TCCTCCCCAGGGTCCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAG 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1182 TCCTCCCCAGGGTCCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAG 1241

Qy        348 TCCTCTAGACTCCTGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAG 407
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1242 TCCTCTAGACTCCTGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAG 1301

Qy        408 TGAAGAGGAGGATACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTA 467
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1302 TGAAGAGGAGGATACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTA 1361

Qy        468 TGCCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAA 527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1362 TGCCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAA 1421

Qy        528 AGAGCCTGTCACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTT 587
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1422 AGAGCCTGTCACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTT 1481

Qy        588 TCCCATGATCTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGAC 647
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1482 TCCCATGATCTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGAC 1541

Qy        648 TGATATGGACCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGA 707
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1542 TGATATGGACCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGA 1601

Qy        708 GGGAAGCCTGATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAG 767
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1602 GGGAAGCCTGATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAG 1661

Qy        768 TATGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGC 827
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1662 TATGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGC 1721

Qy        828 AATAGGG 834
              |||||||
Db       1722 AATAGGG 1728


RESULT 4
HS326L12
LOCUS       HS326L12               37757 bp    DNA     linear   PRI 22-NOV-2001
DEFINITION  Human DNA sequence from clone RP3-326L12 on chromosome Xq27.1-27.3,
            complete sequence.
ACCESSION   AL023279
VERSION     AL023279.2  GI:17065907
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 37757)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 25, 2001 this sequence version replaced gi:3334694.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP3-326L12 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP3-326L12 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP3-326L12 is at 1 in this sequence. The
            true left end of clone RP6-232G24 is at 37658 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .37757
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704L12326"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q27.1-27.3"
                     /clone="RP3-326L12"
                     /clone_lib="RPCI-3"
     repeat_region   3541. .3568
                     /note="14 copies 2 mer gt 89% conserved"
     repeat_region   4365. .4694
                     /note="L1MD repeat: matches 44. .380 of consensus"
     repeat_region   5695. .5739
                     /note="L1PB3 repeat: matches 5824. .5868 of consensus"
     repeat_region   5988. .7855
                     /note="L1MA4A repeat: matches 4434. .6299 of consensus"
     repeat_region   12038. .12372
                     /note="LTR19B repeat: matches 195. .522 of consensus"
     repeat_region   12382. .12605
                     /note="HUERS-P3b repeat: matches 6666. .6893 of consensus"
     repeat_region   12921. .13635
                     /note="HUERS-P3b repeat: matches 5937. .6666 of consensus"
     repeat_region   13949. .15791
                     /note="HUERS-P3b repeat: matches 4170. .5937 of consensus"
     repeat_region   14518. .16363
                     /note="LTR25-internal repeat: matches 3741. .5507 of
                     consensus"
     misc_feature    16440. .16977
                     /note="match: GSS: Em:AQ582187"
     misc_feature    16846. .17124
                     /note="match: GSS: Em:AQ080619"
     misc_feature    16877. .17284
                     /note="match: GSS: Em:AQ191059"
     misc_feature    16942. .17277
                     /note="match: GSS: Em:AQ091802"
```

```
     repeat_region   17783. .18359
                     /note="LTR19B repeat: matches 1. .580 of consensus"
     repeat_region   24178. .24397
                     /note="110 copies 2 mer gg 55% conserved"
     repeat_region   28844. .29294
                     /note="MLT2A repeat: matches 1. .453 of consensus"
     repeat_region   29295. .29743
                     /note="HERVL repeat: matches 1. .452 of consensus"
     repeat_region   29742. .30149
                     /note="HERVL repeat: matches 726. .1133 of consensus"
     repeat_region   30454. .34882
                     /note="HERVL repeat: matches 1133. .5557 of consensus"
     repeat_region   34888. .34960
                     /note="MLT2A repeat: matches 381. .453 of consensus"
     repeat_region   36421. .36560
                     /note="70 copies 2 mer ta 75% conserved"
     misc_feature    37525. .37754
                     /note="match: GSS: Em:AQ503756"
     misc_feature    37525. .37718
                     /note="match: GSS: Em:AQ238088"
BASE COUNT    11703 a   7060 c   7919 g  11075 t
ORIGIN

  Query Match             53.2%;  Score 554;  DB 9;  Length 37757;
  Best Local Similarity   73.5%;  Pred. No. 1e-152;
  Matches  766;  Conservative    0;  Mismatches  260;  Indels   16;  Gaps    4;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              |||||| | |||  |||  ||||  ||  || |||||||||||||||||  | |
Db      25743 ATGCCTGTGTTTTGAAATGTTCCTAGCGGCAACTTTGAGAAAGACTTCCAGAACCAAACT 25802

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              |||| |||| |||||| ||||||||||||||||||| ||| ||||| |||||| | ||||
Db      25803 GTGATAGAGAACTTGGCAGATGCACAGGATTCCACAGAGGAGGAAGAGGAGAAAACCTCC 25862

Qy        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
              |||| |||||||||||||||| |||| ||||||||||||||| ||        | ||| ||
Db      25863 TCCATTTCCTCTTCCTCTTTCTACTTATTATTCCCCTCCTCTTC------CTTCTCTTCT 25916

Qy        181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
              ||  |||| |||||  | ||  ||||| |  |             ||     |
Db      25917 TCCTCCTTCTCCTCCTCTTTCTCCTCTTCCTCCTCCTCCTCCTCATCCTCTACTCTG 25976

Qy        241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300
              || | || |||          |              ||    || ||| ||| |  ||     |
Db      25977 ATTCTTGGTGCTCATGAGGAGGAGGAGGAGTAGCTGTCTGCTGGGATGCTGCC------T 26030

Qy        301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360
              | |||| |||   ||||  || ||||||| ||| |||||| |||| || | ||   || |
Db      26031 CTTCCCCAGAATCCTCCTGAGAGTCCTCCCCAGGGTCCTCTCCAGGGTTCCCTGTCCTGC 26090

Qy        361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAGAGGAGGAT 420
              | ||| ||| || | |  |  |       || |||     ||||   |||| || |||
Db      26091 TCCTCCTCCACTTTATGAAGCCCATTCAATGAAGAGCCTAGCAGCCAATAAGATGAAGAT 26150

Qy        421 ACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAA 480
              |||  ||| |||||   |||||||| || ||  |||||  ||  ||| | |||||||||
Db      26151 ACAAGTACCTGGCACAGCTTGCCAGAGAGCGAGCCCTTGTTCACTTATACACTGGATGAA 26210

Qy        481 AAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACA 540
              |||||||   |||||||||||||||||||||||||||||||| |||||||||| |||||
Db      26211 AAGGTGGACAAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAGCAAAAGAGCCTCTCACA 26270

Qy        541 AAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTC 600
              |  |||||||||| || || |||||||| |  || | || ||| |||||  |||||||||
Db      26271 AGAGCAGAGATGCAGATGAATGTCATCAACACATACACGGGCTACTTTCCTATGATCTTC 26330

Qy        601 GGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCC 660
              |||||||||  |||||||||||||  |  ||||||||||  |||||| ||  ||||||||
Db      26331 AGGAAAGCCCGTGAGTTCATAGAGATTCTTTTTGGCATTTCCCTGACAGAAGTGGACCCC 26390

Qy        661 GACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATT 720
              |   |||| | |||| |||||| | |||||||||||||||||| |||||| || |||| |

Db      26391 G---ACCATTTCTATGTCTTTGTAAACACATTAGACCTCACCTGTGAGGGGAGTCTGAGT 26447

Qy        721 GATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATA 780
              ||||| |||||||||||| ||||| | |||||||||||||||| ||| ||||||||||
Db      26448 GATGAGCAGGGCATGCCCCAGAACCGCCTCCTGATTCTTATTCTGAGTGTGATCTTCATA 26507

Qy        781 AAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGT 840
              ||||||| |||||   | |||||||||||||||||||| ||| ||||||||||||||||
Db      26508 AAGGGCAACTGTGCATCTGAGGAGGTCATCTGGGAAGTGCTGAATGCAATAGGGGTGTGT 26567

Qy        841 GCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTG 900
              |||  |||||||| ||  | ||| |||| |||||  ||||  |||||| |  ||| |||
Db      26568 GCTCAGAGGGAGCATTTCGTCTATCGGGAGCCCAGGGAGCTTGTCACTAAAGTTTGAGTG 26627

Qy        901 CAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAATTT 960
              ||| || | |||||||||||  ||  | |||||||  |||||||||||||||||||||||
Db      26628 CAGGGACATTACCTGGAGTAGCGACATGTGCCCAGTAGTGCTCCTCCACGTTATGAATTC 26687

Qy        961 TTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGA-GAAGTCTTAGAGTTTTTATCCA 1019
              |||||||||| |||||||||||||  | ||||||| |||       ||| ||||   ||
Db      26688 CTGTGGGGTCCTAGAGCCCATTCAGAAACTAGCAAGAGGAAAGTAGTGAGATTTTGGTCA 26747

Qy       1020 AGCTATCCAGTATCATCCCTAG 1041
              |  |  || || | || ||||
Db      26748 TGTGAAACAATACCGTCTCTAG 26769


RESULT 5
HS406C18/c
LOCUS       HS406C18              116210 bp    DNA     linear   PRI 05-MAR-2003
DEFINITION  Human DNA sequence from clone RP3-406C18 on chromosome Xq27.1-27.3,
            complete sequence.
ACCESSION   AL023773
VERSION     AL023773.1  GI:3449129
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 116210)
  AUTHORS   Pearce,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 21, 1998 this sequence version replaced gi:3288039.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
```

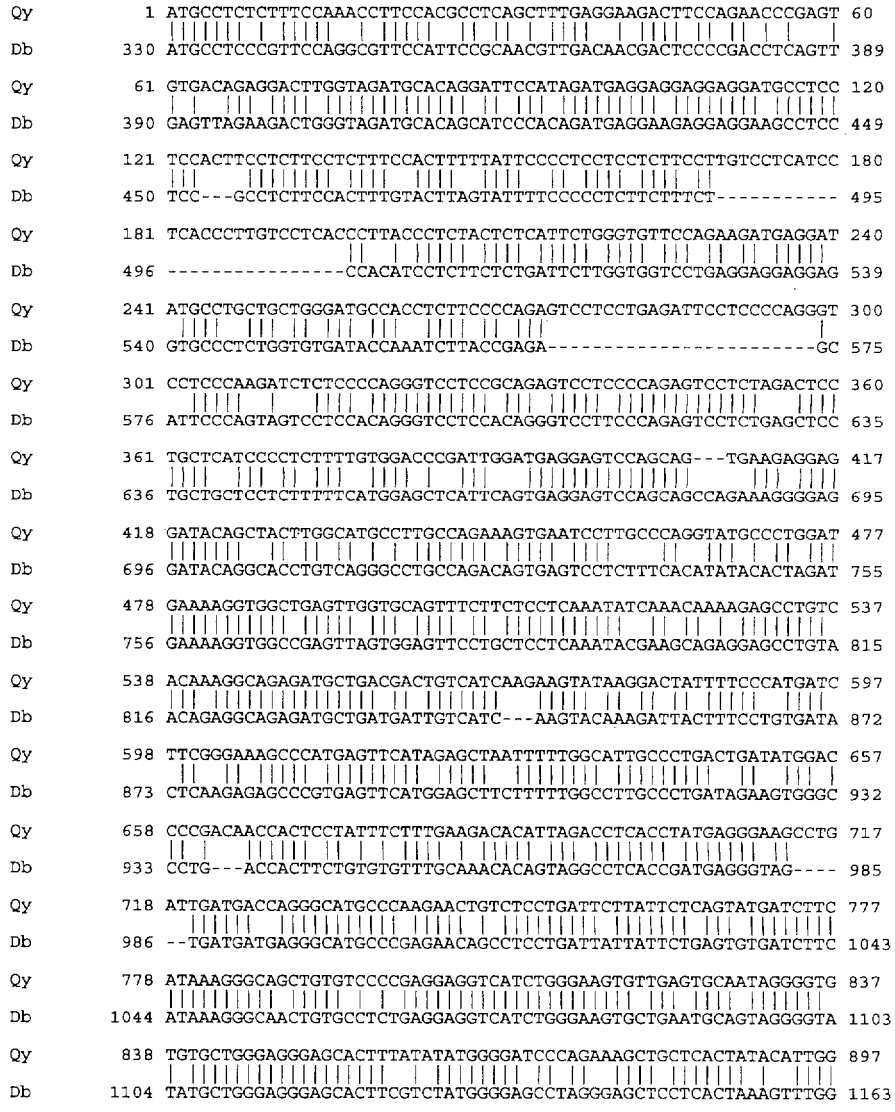
```
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP3-406C18 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES              Location/Qualifiers
     source           1. .116210
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="RZPD:RPCIP704C18406"
                      /db_xref="taxon:9606"
                      /chromosome="X"
                      /map="q27.1-27.3"
                      /clone="RP3-406C18"
                      /clone_lib="RPCI-3"
BASE COUNT    35590 a  22206 c  22413 g  36001 t
ORIGIN

  Query Match             51.3%;  Score 534.4;  DB 9;  Length 116210;
  Best Local Similarity   75.3%;  Pred. No. 7.9e-147;
  Matches  786;  Conservative    0;  Mismatches  191;  Indels   67;  Gaps    7;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              ||||||||||||||| | ||||  || || ||||| ||||||| |||| ||||    |
Db      99751 ATGCCTCTCTTTCCAAGCGTTCCATTCCCCAACTTTGCGGAAGACTCCCAGTACCCAGTT 99692

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              || | |||||| | | |||||||||||||| ||| ||||||||| ||||||||  |   ||
Db      99691 GTAATAGAGGATTGGATAGATGCACAGGATCCCACAGATGAGGAAGAGGAGGAAG---CC 99635

Qy        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
              |||| |||||||||  ||||| ||||  ||||  ||||||| || ||||
Db      99634 TCCATTTCCTCTTCTACTTTCTACTTAGTATTTTCCTCCTCTTCCTCCT---------- 99586

Qy        181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
                          ||| | ||||| |||| ||||| |||| || || || |||||
Db      99585 ----------------CCTCATCCTCTTCTCTGATTCTTGGTGGCCCTGAGGAGGAGGAG 99542

Qy        241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300
              |||    |||||||||||| ||  |||||  |||||  ||||  ||| ||||||  ||||||
Db      99541 GTGCACTCTGCTGGGATGTCAAGTCTTCTCCAGAGCACTCCCAAGAGTCCTCCACAGGGT 99482

Qy        301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360
              ||| ||                  ||  ||||||||||| ||||||    ||||
Db      99481 CCTTCC------------------------CAATGTCCTCCCCAGAATCCTCTGAGCTCC 99446

Qy        361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAGAGGAGGAT 420
              ||||| ||| || |||  || | || | || |||| ||||||||||  || | |||||
Db      99445 TGCTCTTCCTCTGTTTCATGCAGCCCAATGAATGAAGAGTCCAGCAGCCAAAAAGAGGAG 99386

Qy        421 ACAGCT--ACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATG 478
                 ||   || || | || | ||||||| ||||| | |||    ||  ||| | || ||||
Db      99385 GATGCAGGACCTGTCCTGGCCTGCCAGACAGTGAGTTCTTTTTCACATATACACTAGATG 99326

Qy        479 AAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCA 538
              |||||||||| ||||  ||| |||| || ||||||||||||||||||| ||||||  | |
Db      99325 AAAAGGTGGCCGAGTGAGTGGAGTTCCTGCTCCTCAAATATCAAACAAAGGAGCCTATAA 99266

Qy        539 CAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCT 598
              || || |||||||||||||| ||||||||  ||||| || ||||| |||||   ||||
Db      99265 CAGAGTCAGAGATGCTGACGATTGTCATCACCAAGTACAAAGACTACTTTCCTGTGATAC 99206

Qy        599 TCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACC 658
              ||| |||||||  ||||||||||||||||  |||||||| || ||||||  ||  ||| ||
Db      99205 TCGAGAAAGCCTGTGAGTTCATAGAGCTTCTTTTTGGCCTTTCCCTGATAGAAGTGGGCC 99146

Qy        659 CCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGA 718
              | |   |||| | |||| | |||| | ||||| |||||||| || |||||||  ||
Db      99145 CTG---ACCATTTCTATGTGTTTGCAAACACAGTAGACCTCGCCGATGAGGGTAG----- 99094

Qy        719 TTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCA 778
              |||||||  ||||| |||  ||||||| ||||||||||||||||| ||| |||||||||
Db      99093 -TGATGACACGGGCAGGCCTGAGAACTGCCTCCTGATTCTTATTCTGAGTGTGATCTTCA 99035

Qy        779 TAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGT 838
              ||| |||||||||| ||| |||||||||||||||||||| |||  ||||||||||||||
Db      99034 TAACGGGCAGCTGTGCCCCTGAGGAGGTCATCTGGGAAGTGCTGAAGGCAATAGGGGTGT 98975

Qy        839 GTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGGG 898
              |||||||||||||||||  | |||||||  |||||  |||  |||||||  |  |||||
Db      98974 ATGCTGGGAGGGAGCACTTCGTCTATGGGGAGCCCAGGGAGCTCCTCACTAAAGTTTGGG 98915

Qy        899 TGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAAT 958
              |||||  | | ||||||||||  |||||||||||||||| ||||||||||||||||||
Db      98914 TGCAGGAACATTACCTGGAGTATTGGGAGGTGCCCAACAGTTCTCCTCCACGTTATGAAT 98855

Qy        959 TTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAG--AGAAGTCTTAGAGTTTTTAT 1016
              |  |||||||||||||||||||||||||    || ||||    ||||  ||||||||||
Db      98854 TCCTGTGGGGTCCAAGAGCCCATTCAGAAAGCACCAAGCAGAAAGTACTAGAGTTTTTGG 98795

Qy       1017 CCAAGCTATCCAGTATCATCCCTA 1040
              |||||||| || ||  |||||||
Db      98794 CCAAGCTATACAATAGTATCCCTA 98771


RESULT 6
AF333706
LOCUS       AF333706                 533 bp    DNA     linear   PRI 01-FEB-2001
DEFINITION  Homo sapiens MAGE family testis and tumor-specific protein (MAGEC3)
            gene, partial cds.
ACCESSION   AF333706
VERSION     AF333706.1  GI:12642822
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   Lucas,S., De Plaen,E. and Boon,T.
  TITLE     MAGE-B5, MAGE-B6, MAGE-C2, and MAGE-C3: four new members of the
            MAGE family with tumor-specific expression
  JOURNAL   Int. J. Cancer 87 (1), 55-60 (2000)
  MEDLINE   20321428
   PUBMED   10861452
REFERENCE   2  (bases 1 to 533)
  AUTHORS   Lucas,S., De Plaen,E. and Boon,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2001) Ludwig Institute for Cancer Research,
            Catholic University of Louvain, Avenue Hippocrate, 74, Brussels
            1200, Belgium
FEATURES              Location/Qualifiers
     source           1. .533
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="X"
                      /map="Xq27.1-q27.3"
                      /tissue_type="whole blood"
     gene             <1. .>533
                      /gene="MAGEC3"
     CDS              <1. .>533
                      /gene="MAGEC3"
                      /codon_start=3
                      /product="MAGE family testis and tumor-specific protein"
                      /protein_id="AAK00358.1"
                      /db_xref="GI:12642823"
                      /translation="SSPLPSTLILGVPEDEDMPAAGMPPLPQSPPEIPPQGPPKISPQ
                      GPPQSPPQSPLDSCSSPLLWTRLDEESSSEEEDTATWHALPESESLPRYALDEKVAEL
                      VQFLLLKYQTKEPVTKAEMLTTVIKKYKDYFPMIFGKAHEFIELIFGIALTDMDPDNH
                      SYFFEDTLDLTYEGSLI"
BASE COUNT      126 a    151 c    119 g    137 t
ORIGIN

  Query Match             51.2%;  Score 533;  DB 9;  Length 533;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-147;
```

```
  Matches  533;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        188 TGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGATATGCCTG 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGATATGCCTG 60

Qy        248 CTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGTCCTCCCA 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGTCCTCCCA 120

Qy        308 AGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCAT 367
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCAT 180

Qy        368 CCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAGAGGAGGATACAGCTA 427
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAGAGGAGGATACAGCTA 240

Qy        428 CTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGG 487
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGG 300

Qy        488 CTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAG 547
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAG 360

Qy        548 AGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAG 607
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAG 420

Qy        608 CCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACC 667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACC 480

Qy        668 ACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATT 533


RESULT 7
BC005891
LOCUS       BC005891               1734 bp    mRNA    linear   PRI 06-JUN-2003
DEFINITION  Homo sapiens melanoma antigen, family E, 1, cancer/testis specific,
            mRNA (cDNA clone MGC:871 IMAGE:3050294), complete cds.
ACCESSION   BC005891
VERSION     BC005891.1  GI:13543465
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1734)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1734)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 3 Row: f Column: 20
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 7706604.
FEATURES             Location/Qualifiers
     source          1. .1734
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:871 IMAGE:3050294"
                     /tissue_type="Skin, melanotic melanoma."
                     /clone_lib="NIH_MGC_20"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1. .1734
                     /gene="MAGEE1"
                     /note="synonyms: MAGEC2, CT10, MAGE-C2, MGC13377, HCA587"
                     /db_xref="LocusID:51438"
     CDS             89. .1210
                     /codon_start=1
                     /product="melanoma antigen, family E, 1, cancer/testis
                     specific"
                     /protein_id="AAH05891.1"
                     /db_xref="GI:13543466"
                     /db_xref="LocusID:51438"
                     /translation="MPPVPGVPFRNVDNDSPTSVELEDWVDAQHPTDEEEEEASSASS
                     TLYLVFSPSSSFSTSSSLILGGPEEEEVPSGVIPNLTESIPSSPPQGPPQGPSQSPLSS
                     CCSSFSWSSFSEESSSQKGEDTGTCQGLPDSESSFTYTLDEKVAELVEFLLLKYEAEE
                     PVTEAEMLMIVIKYKDYFPVILKRAREFMELLFGLALIEVGPDHFCVFANTVGLTDEG
                     SDDEGMPENSLLIIILSVIFIKGNCASEEVIWEVLNAVGVYAGREHFVYGEPRELLTK
                     VWVQGHYLEYREVPHSSPPYYEFLWGPRAHSESIKKKVLEFLAKLNNTVPSSFPSWYK
                     DALKDVEERVQATIDTADDATVMASESLSVMSSNVSFSE"
BASE COUNT      431 a    371 c    424 g    508 t
ORIGIN

  Query Match           46.7%;  Score 486.4;  DB 9;  Length 1734;
  Best Local Similarity  73.0%;  Pred. No. 8.1e-133;
  Matches  764;  Conservative    0;  Mismatches  211;  Indels   71;  Gaps    8;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              ||||||| | |||||  | |||||  || || | ||||  | |||| || || | |   |
Db         89 ATGCCTCCCGTTCCAGGCGTTCCATTCCGCAACGTTGACAACGACTCCCCGACCTCAGTT 148

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              | |  ||| |||| |||||||||||||| || ||| |||||||| |||||||| ||||||
Db        149 GAGTTAGAAGACTGGGTAGATGCACAGCATCCCACAGATGAGGAAGAGGAGGAAGCCTCC 208

Qy        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
              |||    |||||||| ||||  ||||  ||||  || |||| |||| ||
Db        209 TCC---GCCTCTTCCACTTTGTACTTAGTATTTTCCCCCTCTTCTTTCT---------- 254
```



RESULT 8
AR243321
LOCUS AR243321 1983 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 18 from patent US 6475783.
ACCESSION AR243321
VERSION AR243321.1 GI:27290517
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1983)
AUTHORS Lucas,S., De Smet,C. and Boon-Falleur,T.
TITLE Isolated nucleic acid molecule coding for tumor rejection antigen precursors MAGE-C1 and MAGE-C2 and uses thereof
JOURNAL Patent: US 6475783-A 18 05-NOV-2002;
FEATURES Location/Qualifiers
source 1. .1983
/organism="unknown"
BASE COUNT 497 a 428 c 512 g 546 t
ORIGIN

Query Match 46.7%; Score 486.4; DB 6; Length 1983;
Best Local Similarity 73.0%; Pred. No. 8.2e-133;
Matches 764; Conservative 0; Mismatches 211; Indels 71; Gaps 8;



```
Qy        898 GTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCCAACAGTGCTCCTCCACGTTATGAA 957
              |||||| || | |||||||||||| |||||||||||| ||||| ||||||||  |||||||
Db       1164 GTGCAGGGACATTACCTGGAGTATCGGGAGGTGCCCCACAGTTCTCCTCCATATTATGAA 1223

Qy        958 TTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGA--GAAGTCTTAGAGTTTTTA 1015
              ||  ||||||||||||||||||||||||||    || |||||   ||||  |||||||||||
Db       1224 TTCCTGTGGGGTCCAAGAGCCCATTCAGAAAGCATCAAGAAGAAAGTACTAGAGTTTTTA 1283

Qy       1016 TCCAAGCTATCCAGTATCATCCCTAG 1041
              |||||||   ||  |   | |||||
Db       1284 GCCAAGCTGAACAACACTGTTCCTAG 1309


RESULT 9
BD106883
LOCUS       BD106883               1983 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Isolated nucleic acid molecule coding for tumor rejection antigen
            precursors MAGE-C1 and MAGE-C2 and uses thereof.
ACCESSION   BD106883
VERSION     BD106883.1  GI:23201701
KEYWORDS    JP 2002503096-A/16.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 1983)
  AUTHORS   Lucas,S., Smet,C.D. and Falleur,T.B.
  TITLE     Isolated nucleic acid molecule coding for tumor rejection antigen
            precursors MAGE-C1 and MAGE-C2 and uses thereof
  JOURNAL   Patent: JP 2002503096-A 16 29-JAN-2002;
            LUDWIG INSTITUTE FOR CANCER RESEARCH
COMMENT     PN   JP 2002503096-A/16
            PD   29-JAN-2002
            PF   24-APR-1998 JP 1998547266
            PR   25-APR-1997 US    08/845528
            PI   SOPHIE LUCAS,CHARLES DE SMET,THIERRY BOON FALLEUR PC
            C07H21/04,A61K38/00,A61K39/00,A61K39/12,G01N33/574,C07K5/00, PC
            C07K7/00,
            PC   C07K16/00,C07K17/00
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers.
FEATURES              Location/Qualifiers
     source           1. .1983
                      /organism="synthetic construct"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:32630"
BASE COUNT      497 a    428 c    512 g    546 t
ORIGIN

  Query Match             46.7%;  Score 486.4;  DB 6;  Length 1983;
  Best Local Similarity   73.0%;  Pred. No. 8.2e-133;
  Matches  764;  Conservative    0;  Mismatches  211;  Indels   71;  Gaps    8;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              ||||||| | |||||  | |||||  || || | |||  | |||| || || | |   |
Db        330 ATGCCTCCCGTTCCAGGCGTTCCATTCCGCAACGTTGACAACGACTCCCCGACCTCAGTT 389

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              | |  ||| |||| |||||||||||| || ||| |||||||| |||||||| ||||||
Db        390 GAGTTAGAAGACTGGGTAGATGCACAGCATCCCACAGATGAGGAAGAGGAGGAAGCCTCC 449

Qy        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
              |||    |||||||| ||||  ||||  ||||  || |||| |||| ||
Db        450 TCC---GCCTCTTCCACTTTGTACTTAGTATTTTCCCCCTCTTCTTTCT---------- 495

Qy        181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
                          ||  | ||||| |||| ||||| |||| ||| || || |||||
Db        496 ----------------CCACATCCTCTTCTCTGATTCTTGGTGGTCCTGAGGAGGAGGAG 539

Qy        241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300
              ||||  ||| || ||| |||  |||| || |||                        |
Db        540 GTGCCCTCTGGTGTGATACCAAATCTTACCGAGA-----------------------GC 575

Qy        301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360
              |||||  |   |||| |||||||||||| ||| ||||| ||||||||||||    ||||
Db        576 ATTCCCAGTAGTCCTCCACAGGGTCCTCCACAGGGTCCTTCCCAGAGTCCTCTGAGCTCC 635

Qy        361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAG 417
              |||  ||| || |||  |||| |  |||   ||||||||||||||     ||| ||||
Db        636 TGCTGCTCCTCTTTTTCATGGAGCTCATTCAGTGAGGAGTCCAGCAGCCAGAAAGGGGAG 695

Qy        418 GATACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGAT 477
              |||||||  || || || | | ||||||| ||||| ||||    ||  ||| | || |||
Db        696 GATACAGGCACCTGTCAGGGCCTGCCAGACAGTGAGTCCTCTTTCACATATACACTAGAT 755

Qy        478 GAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTC 537
              |||||||||| ||||| ||| |||| || |||||||||| || || | ||||||||
Db        756 GAAAAGGTGGCCGAGTTAGTGGAGTTCCTGCTCCTCAAATACGAAGCAGAGGAGCCTGTA 815

Qy        538 ACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATC 597
              ||| |||||||||||||| || |||||||    ||||| || || || |||||   ||||
Db        816 ACAGAGGCAGAGATGCTGATGATTGTCATC---AAGTACAAAGATTACTTTCCTGTGATA 872

Qy        598 TTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGAC 657
              ||  || ||||| |||||||| |||||  |||||||| |||||||| ||  ||| |
Db        873 CTCAAGAGAGCCCGTGAGTTCATGGAGCTTCTTTTTGGCCTTGCCCTGATAGAAGTGGGC 932

Qy        658 CCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTG 717
              ||  |  |||||| || | | |||| | ||||| ||| |||||||| ||||||| ||
Db        933 CCTG---ACCACTTCTGTGTGTTTGCAAACACAGTAGGCCTCACCGATGAGGGTAG---- 985

Qy        718 ATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTC 777
              |||||  |||||||||| ||||| | |||||||| |||||| ||| |||||||||||
Db        986 --TGATGATGAGGGCATGCCCGAGAACAGCCTCCTGATTATTATTCTGAGTGTGATCTTC 1043

Qy        778 ATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTG 837
              ||||||||| ||||| | | |||||||||||||||||||| ||| |||| |||||||
Db       1044 ATAAAGGGCAACTGTGCCTCTGAGGAGGTCATCTGGGAAGTGCTGAATGCAGTAGGGGTA 1103

Qy        838 TGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGG 897
              | |||||||||||||||||| | |||||||  || ||  |||| |||||||  |  ||||
Db       1104 TATGCTGGGAGGGAGCACTTCGTCTATGGGGAGCCTAGGGAGCTCCTCACTAAAGTTTGG 1163

Qy        898 GTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAA 957
              |||||| || | |||||||||||| |||||||||||| ||||| ||||||||  |||||||
Db       1164 GTGCAGGGACATTACCTGGAGTATCGGGAGGTGCCCCACAGTTCTCCTCCATATTATGAA 1223

Qy        958 TTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGA--GAAGTCTTAGAGTTTTTA 1015
              ||  ||||||||||||||||||||||||||    || |||||   ||||  |||||||||||
Db       1224 TTCCTGTGGGGTCCAAGAGCCCATTCAGAAAGCATCAAGAAGAAAGTACTAGAGTTTTTA 1283

Qy       1016 TCCAAGCTATCCAGTATCATCCCTAG 1041
              |||||||   ||  |   | |||||
Db       1284 GCCAAGCTGAACAACACTGTTCCTAG 1309


RESULT 10
AF151378
LOCUS       AF151378               1983 bp    mRNA    linear   PRI 17-JUL-2002
DEFINITION  Homo sapiens hepatocellular cancer antigen 587 (HCA587) mRNA,
            complete cds.
ACCESSION   AF151378
VERSION     AF151378.1  GI:7108900
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1983)
  AUTHORS   Wang,Y., Han,K.J., Pang,X.W., Vaughan,H.A., Qu,W., Dong,X.Y.,
            Peng,J.R., Zhao,H.T., Rui,J.A., Leng,X.S., Cebon,J., Burgess,A.W.
            and Chen,W.F.
  TITLE     Large scale identification of human hepatocellular
            carcinoma-associated antigens by autoantibodies
  JOURNAL   J. Immunol. 169 (2), 1102-1109 (2002)
  MEDLINE   22092308
```

```
   PUBMED   12097419
REFERENCE   2  (bases 1 to 1983)
  AUTHORS   Qu,W., Wang,Y., Han,K.J. and Chen,W.F.
  TITLE     A new MAGE family gene identified in liver cancer
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1983)
  AUTHORS   Qu,W., Wang,Y., Han,K.J. and Chen,W.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-1999) Immunology, Beijing Medical University, #38
            Xueyuan Road, Beijing 100083, P. R. China
FEATURES             Location/Qualifiers
     source          1. .1983
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /note="isolated from homologous serum of a hepatoma
                     patient"
     gene            1. .1983
                     /gene="HCA587"
                     /note="MAGE gene family"
     CDS             330. .1451
                     /gene="HCA587"
                     /codon_start=1
                     /product="hepatocellular cancer antigen 587"
                     /protein_id="AAF36533.1"
                     /db_xref="GI:7108901"
                     /translation="MPPVPGVPFRNVDNDSPTSVELEDWVDAQHPTDEEEEEASSASS
                     TLYLVFSPSSFSTSSSLILGGPEEEEVPSGVIPNLTESIPSSPPQGPPQGPSQSPLSS
                     CCSSFSWSSFSEESSSQKGEDTGTCQGLPDSESSFTYTLDEKVAELVEFLLLKYEAEE
                     PVTEAEMLMIVIKYKDYFPVILKRAREFMELLFGLALIEVGPDHFCVFANTVGLTDEG
                     SDDEGMPENSLLIIILSVIFIKGNCASEEVIWEVLNAVGVYAGREHFVYGEPRELLTK
                     VWVQGHYLEYREVPHSSPPYYEFLWGPRAHSESIKKKVLEFLAKLNNTVPSSFPSWYK
                     DALKDVEERVQATIDTADDATVMASESLSVMSSNVSFSE"
BASE COUNT      497 a    428 c    512 g    546 t
ORIGIN

  Query Match           46.7%;  Score 486.4;  DB 9;  Length 1983;
  Best Local Similarity  73.0%;  Pred. No. 8.2e-133;
  Matches  764;  Conservative    0;  Mismatches  211;  Indels   71;  Gaps    8;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              ||||||| | |||||  | |||||  || || | ||||  | |||| || || | |   |
Db        330 ATGCCTCCCGTTCCAGGCGTTCCATTCCGCAACGTTGACAACGACTCCCCGACCTCAGTT 389

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              | |  ||| |||| |||||||||||||| || ||| |||||||| |||||||| ||||||
Db        390 GAGTTAGAAGACTGGGTAGATGCACAGCATCCCACAGATGAGGAAGAGGAGGAAGCCTCC 449

Qy        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
              |||    |||||||| ||||  ||||  ||||  || |||| |||| ||
Db        450 TCC---GCCTCTTCCACTTTGTACTTAGTATTTTCCCCCTCTTCTTTCT---------- 495

Qy        181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
                          ||  | |||||  |||| |||||  |||| |||  || || |||||
Db        496 ----------------CCACATCCTCTTCTCTGATTCTTGGTGGTCCTGAGGAGGAGGAG 539

Qy        241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300
              ||||  |||  || ||| |||  |||| || |||                        |
Db        540 GTGCCCTCTGGTGTGATACCAAATCTTACCGAGA-----------------------GC 575

Qy        301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360
                |||||  |   |||| |||||||||||| ||| ||||| ||||||||||||||   ||||
Db        576 ATTCCCAGTAGTCCTCCACAGGGTCCTCCACAGGGTCCTTCCCAGAGTCCTCTGAGCTCC 635

Qy        361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAG 417
              ||||  ||| || |||  |||| |  |||   |||||||||||||||     ||| ||||
Db        636 TGCTGCTCCTCTTTTTCATGGAGCTCATTCAGTGAGGAGTCCAGCAGCCAGAAAGGGGAG 695

Qy        418 GATACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGAT 477
              |||||||  || || || | | ||||||| ||||| ||||    ||  ||| | || |||
Db        696 GATACAGGCACCTGTCAGGGCCTGCCAGACAGTGAGTCCTCTTTCACATATACACTAGAT 755

Qy        478 GAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTC 537
              ||||||||||| ||||| ||| |||| || |||||||||||  || || | ||||||||
Db        756 GAAAAGGTGGCCGAGTTAGTGGAGTTCCTGCTCCTCAAATACGAAGCAGAGGAGCCTGTA 815

Qy        538 ACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATC 597
              ||| |||||||||||||| || |||||||   ||||| || || || |||||  ||||
Db        816 ACAGAGGCAGAGATGCTGATGATTGTCATC---AAGTACAAAGATTACTTTCCTGTGATA 872

Qy        598 TTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGAC 657
                ||  || ||||| |||||||| |||||  |||||||| ||||||||  ||  ||| |
Db        873 CTCAAGAGAGCCCGTGAGTTCATGGAGCTTCTTTTTGGCCTTGCCCTGATAGAAGTGGGC 932

Qy        658 CCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTG 717
              || |   |||||| || | | |||| | ||||| ||| ||||||| |||||||  ||
Db        933 CCTG---ACCACTTCTGTGTGTTTGCAAACACAGTAGGCCTCACCGATGAGGGTAG---- 985

Qy        718 ATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTC 777
                ||||||  |||||||||||| ||||| | |||||||| ||||||| ||| ||||||||
Db        986 --TGATGATGAGGGCATGCCCGAGAACAGCCTCCTGATTATTATTCTGAGTGTGATCTTC 1043

Qy        778 ATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTG 837
              |||||||||| ||||| | | |||||||||||||||||||| ||| |||| |||||||
Db       1044 ATAAAGGGCAACTGTGCCTCTGAGGAGGTCATCTGGGAAGTGCTGAATGCAGTAGGGGTA 1103

Qy        838 TGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGG 897
              | |||||||||||||||||  | |||||||| || ||  |||| |||||||  |  ||||
Db       1104 TATGCTGGGAGGGAGCACTTCGTCTATGGGGAGCCTAGGGAGCTCCTCACTAAAGTTTGG 1163

Qy        898 GTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAA 957
              |||||| || | |||||||||||| |||||||||||| ||||| ||||||||  |||||||
Db       1164 GTGCAGGGACATTACCTGGAGTATCGGGAGGTGCCCCACAGTTCTCCTCCATATTATGAA 1223

Qy        958 TTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGA--GAAGTCTTAGAGTTTTTA 1015
              ||  |||||||||||||||||||||||||  || |||||   ||||  ||||||||||||
Db       1224 TTCCTGTGGGGTCCAAGAGCCCATTCAGAAAGCATCAAGAAGAAAGTACTAGAGTTTTTA 1283

Qy       1016 TCCAAGCTATCCAGTATCATCCCTAG 1041
              |||||||   ||  |   | |||||
Db       1284 GCCAAGCTGAACAACACTGTTCCTAG 1309
```

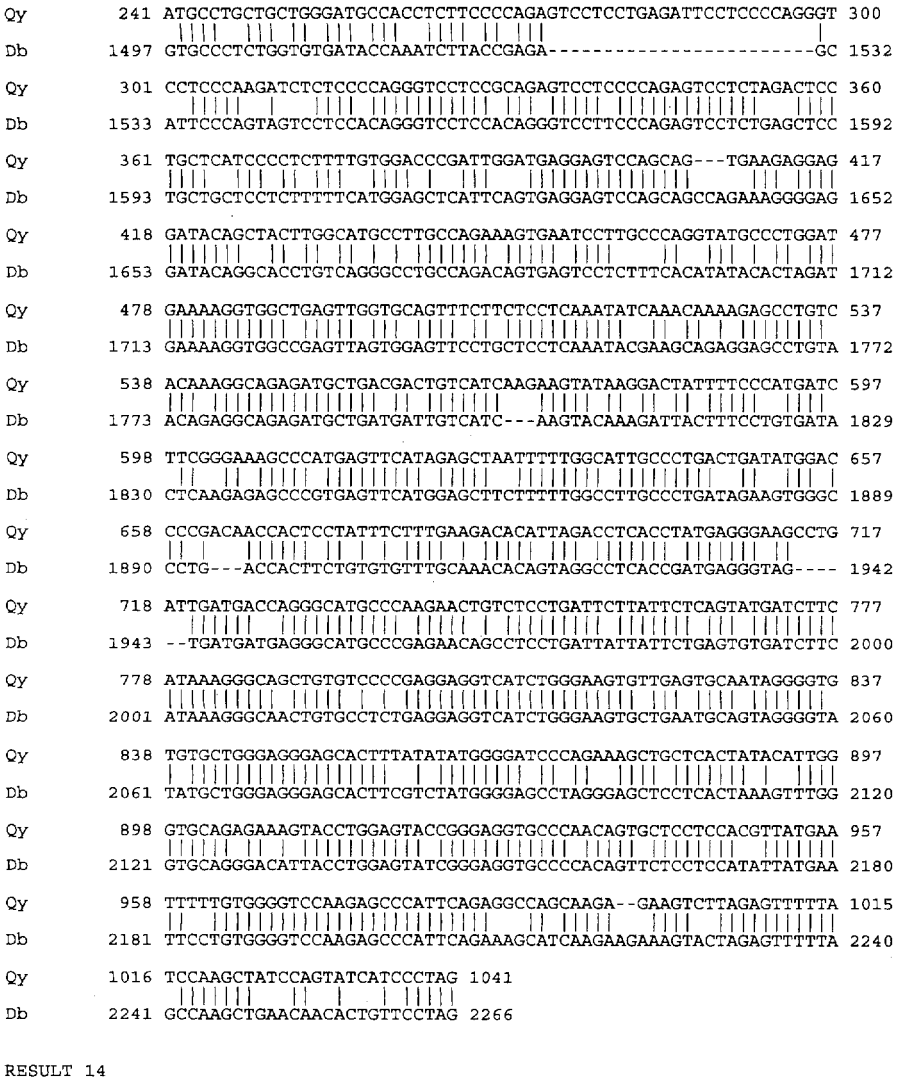
```
RESULT 11
AF196482
LOCUS       AF196482               1983 bp    mRNA    linear   PRI 10-NOV-1999
DEFINITION  Homo sapiens MAGE-C2 (MAGEC2) mRNA, complete cds.
ACCESSION   AF196482
VERSION     AF196482.1  GI:6319213
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1983)
  AUTHORS   Lucas,S., De Plaen,E. and Boon,T.
  TITLE     MAGE-B5, MAGE-B6, MAGE-C2 AND MAGE-C3: Four new members of the mage
            family with tumor-specific expression
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1983)
  AUTHORS   Lucas,S., De Plaen,E. and Boon,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1999) Ludwig Institute for Cancer Research,
            Universite Catholique de Louvain, avenue Hippocrate, 74, Brussels
            1200, Belgium
FEATURES             Location/Qualifiers
     source          1. .1983
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="between DXS1227 and DXS7087"
                     /cell_line="LB373-MEL"
                     /cell_type="melanoma"
     gene            1. .1983
                     /gene="MAGEC2"
```

```
     CDS             330. .1451
                     /gene="MAGEC2"
                     /note="member of the MAGE protein family"
                     /codon_start=1
                     /product="MAGE-C2"
                     /protein_id="AAF07210.1"
                     /db_xref="GI:6319214"
                     /translation="MPPVPGVPFRNVDNDSPTSVELEDWVDAQHPTDEEEEEASSASS
                     TLYLVFSPSSFSTSSSLILGGPEEEEVPSGVIPNLTESIPSSPPQGPPQGPSQSPLSS
                     CCSSFSWSSFSEESSSQKGEDTGTCQGLPDSESSFTYTLDEKVAELVEFLLLKYEAEE
                     PVTEAEMLMIVIKYKDYFPVILKRAREFMELLFGLALIEVGPDHFCVFANTVGLTDEG
                     SDDEGMPENSLLIIILSVIFIKGNCASEEVIWEVLNAVGVYAGREHFVYGEPRELLTK
                     VWVQGHYLEYREVPHSSPPYYEFLWGPRAHSESIKKKVLEFLAKLNNTVPSSFPSWYK
                     DALKDVEERVQATIDTADDATVMASESLSVMSSNVSFSE"
BASE COUNT      497 a    428 c    512 g    546 t
ORIGIN

  Query Match           46.7%;  Score 486.4;  DB 9;  Length 1983;
  Best Local Similarity  73.0%;  Pred. No. 8.2e-133;
  Matches  764;  Conservative    0;  Mismatches  211;  Indels   71;  Gaps    8;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              ||||||| | |||||  | |||||  || || | ||||  | |||| || || | |   |
Db        330 ATGCCTCCCGTTCCAGGCGTTCCATTCCGCAACGTTGACAACGACTCCCCGACCTCAGTT 389

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              | |  ||| |||| |||||||||||||| || ||| ||||||||| |||||||| ||||||
Db        390 GAGTTAGAAGACTGGGTAGATGCACAGCATCCCACAGATGAGGAAGAGGAGGAAGCCTCC 449

Qy        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
              |||    |||||||| ||||  ||||  ||||  || |||| |||| ||
Db        450 TCC---GCCTCTTCCACTTTGTACTTAGTATTTTCCCCCTCTTCTTTCT---------- 495

Qy        181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
                          ||  | ||||| |||| ||||| |||| ||| || || |||||
Db        496 ----------------CCACATCCTCTTCTCTGATTCTTGGTGGTCCTGAGGAGGAGGAG 539

Qy        241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300
              ||||  ||| || ||| |||  |||| || |||                        |
Db        540 GTGCCCTCTGGTGTGATACCAAATCTTACCGAGA-----------------------GC 575

Qy        301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360
              |||||  |   |||| |||||||||||| ||| ||||| ||||||||||||   ||||
Db        576 ATTCCCAGTAGTCCTCCACAGGGTCCTCCACAGGGTCCTTCCCAGAGTCCTCTGAGCTCC 635

Qy        361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAG 417
              |||  ||| || |||  |||| |  |||   |||||||||||||||      ||| ||||
Db        636 TGCTGCTCCTCTTTTTCATGGAGCTCATTCAGTGAGGAGTCCAGCAGCCAGAAAGGGGAG 695

Qy        418 GATACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGAT 477
              |||||||  || || | | |||||||  |||||  ||||     ||  ||| | || |||
Db        696 GATACAGGCACCTGTCAGGGCCTGCCAGACAGTGAGTCCTCTTTCACATATACACTAGAT 755

Qy        478 GAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTC 537
              ||||||||||| ||||| ||| |||| || ||||||||||  || || | ||||||||
Db        756 GAAAAGGTGGCCGAGTTAGTGGAGTTCCTGCTCCTCAAATACGAAGCAGAGGAGCCTGTA 815

Qy        538 ACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATC 597
              ||| |||||||||||||| || |||||||    |||||  ||  || ||  |||||  ||||
Db        816 ACAGAGGCAGAGATGCTGATGATTGTCATC---AAGTACAAAGATTACTTTCCTGTGATA 872

Qy        598 TTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGAC 657
              ||  ||  ||||| ||||||||| |||||  |||||||| |||||||||   ||  ||| |
Db        873 CTCAAGAGAGCCCGTGAGTTCATGGAGCTTCTTTTTGGCCTTGCCCTGATAGAAGTGGGC 932

Qy        658 CCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTG 717
              || |   |||||| || | | |||| | |||||  ||| |||||||  |||||||  ||
Db        933 CCTG---ACCACTTCTGTGTGTTTGCAAACACAGTAGGCCTCACCGATGAGGGTAG---- 985

Qy        718 ATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTC 777
                |||||  |||||||||| |||||  | |||||||||  |||||||  ||| ||||||||
Db        986 --TGATGATGAGGGCATGCCCGAGAACAGCCTCCTGATTATTATTCTGAGTGTGATCTTC 1043

Qy        778 ATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTG 837
              |||||||||| |||||  | | |||||||||||||||||||||||| ||| |||| |||||||
Db       1044 ATAAAGGGCAACTGTGCCTCTGAGGAGGTCATCTGGGAAGTGCTGAATGCAGTAGGGGTA 1103

Qy        838 TGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGG 897
              | |||||||||||||||||  | |||||||| || ||  |||| ||||||| |  ||||
Db       1104 TATGCTGGGAGGGAGCACTTCGTCTATGGGGAGCCTAGGGAGCTCCTCACTAAAGTTTGG 1163

Qy        898 GTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCCAACAGTGCTCCTCCACGTTATGAA 957
              ||||||  || | |||||||||||  |||||||||||  |||||  ||||||||  |||||||
Db       1164 GTGCAGGGACATTACCTGGAGTATCGGGAGGTGCCCCACAGTTCTCCTCCATATTATGAA 1223

Qy        958 TTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGA--GAAGTCTTAGAGTTTTTA 1015
              ||  |||||||||||||||||||||||||    || |||||    ||||  ||||||||||
Db       1224 TTCCTGTGGGGTCCAAGAGCCCATTCAGAAAGCATCAAGAAGAAAGTACTAGAGTTTTTA 1283

Qy       1016 TCCAAGCTATCCAGTATCATCCCTAG 1041
              |||||||   ||  |   | |||||
Db       1284 GCCAAGCTGAACAACACTGTTCCTAG 1309


RESULT 12
AF196483
LOCUS       AF196483               2887 bp    DNA     linear   PRI 10-NOV-1999
DEFINITION  Homo sapiens MAGE-C2 (MAGEC2) gene, complete cds.
ACCESSION   AF196483
VERSION     AF196483.1  GI:6319215
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2887)
  AUTHORS   Lucas,S., De Plaen,E. and Boon,T.
  TITLE     MAGE-B5, MAGE-B6, MAGE-C2 AND MAGE-C3: Four new members of the mage
            family with tumor-specific expression
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2887)
  AUTHORS   Lucas,S., De Plaen,E. and Boon,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1999) Ludwig Institute for Cancer Research,
            Universite Catholique de Louvain, avenue Hippocrate, 74, Brussels
            1200, Belgium
FEATURES             Location/Qualifiers
     source          1. .2887
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="between DXS1227 and DXS7087"
                     /cell_line="LB23"
                     /cell_type="sarcoma"
     mRNA            join(1. .169,819. .912,1221. .2887)
                     /product="MAGE-C2"
     gene            1287. .2408
                     /gene="MAGEC2"
     CDS             1287. .2408
                     /gene="MAGEC2"
                     /note="member of the MAGE protein family"
                     /codon_start=1
                     /product="MAGE-C2"
                     /protein_id="AAF07211.1"
                     /db_xref="GI:6319216"
                     /translation="MPPVPGVPFRNVDNDSPTSVELEDWVDAQHPTDEEEEEASSASS
                     TLYLVFSPSSFSTSSSLILGGPEEEEVPSGVIPNLTESIPSSPPQGPPQGPSQSPLSS
                     CCSSFSWSSFSEESSSQKGEDTGTCQGLPDSESSFTYTLDEKVAELVEFLLLKYEAEE
                     PVTEAEMLMIVIKYKDYFPVILKRAREFMELLFGLALIEVGPDHFCVFANTVGLTDEG
                     SDDEGMPENSLLIIILSVIFIKGNCASEEVIWEVLNAVGVYAGREHFVYGEPRELLTK
                     VWVQGHYLEYREVPHSSPPYYEFLWGPRAHSESIKKKVLEFLAKLNNTVPSSFPSWYK
                     DALKDVEERVQATIDTADDATVMASESLSVMSSNVSFSE"
BASE COUNT      661 a    708 c    794 g    724 t
ORIGIN
```

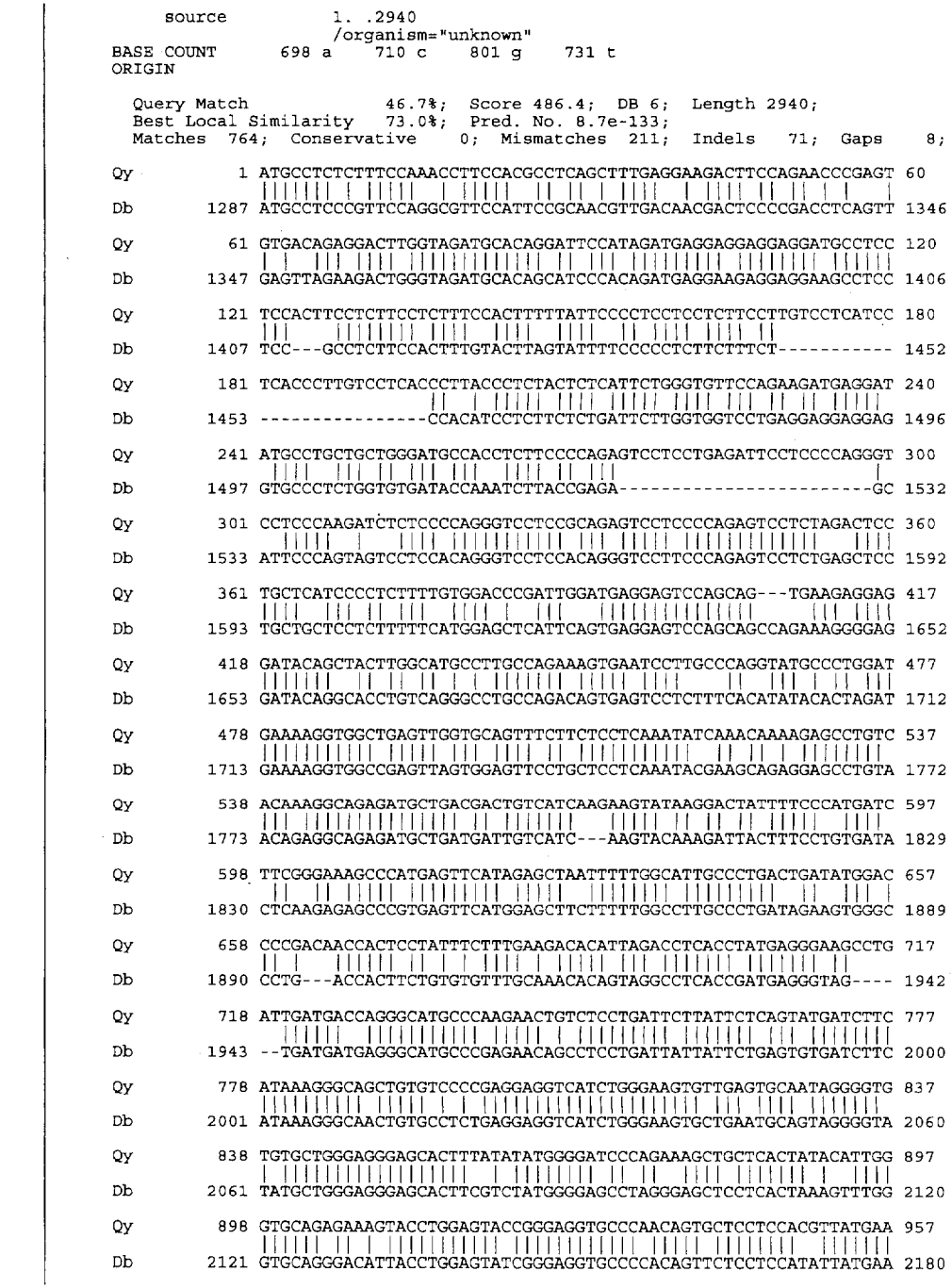
```
  Query Match            46.7%;  Score 486.4;  DB 9;  Length 2887;
  Best Local Similarity  73.0%;  Pred. No. 8.7e-133;
  Matches  764;  Conservative    0;  Mismatches  211;  Indels   71;  Gaps    8;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              ||||||| | |||||  | |||||  || || | ||||  | ||||  || || | |  |
Db       1287 ATGCCTCCCGTTCCAGGCGTTCCATTCCGCAACGTTGACAACGACTCCCCGACCTCAGTT 1346

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              | |  ||| |||| ||||||||||||| || ||| |||||||| ||||||| ||||||
Db       1347 GAGTTAGAAGACTGGGTAGATGCACAGCATCCCACAGATGAGGAAGAGGAGGAAGCCTCC 1406

Qy         121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
               |||    |||||||| ||||  ||||  ||||  || |||| |||| ||
Db       1407 TCC---GCCTCTTCCACTTTGTACTTAGTATTTTCCCCCTCTTCTTTCT---------- 1452

Qy         181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
                         || | ||||| |||| ||||| |||| ||| || || |||||
Db       1453 ----------------CCACATCCTCTTCTCTGATTCTTGGTGGTCCTGAGGAGGAGGAG 1496

Qy         241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300
               ||||  ||| || ||| |||  |||| || |||                        |
Db       1497 GTGCCCTCTGGTGTGATACCAAATCTTACCGAGA-----------------------GC 1532

Qy         301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360
               |||||  |   |||| |||||||||||| ||| ||||| ||||||||||||   ||||
Db       1533 ATTCCCAGTAGTCCTCCACAGGGTCCTCCACAGGGTCCTTCCCAGAGTCCTCTGAGCTCC 1592

Qy         361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAG 417
               ||||  ||| || |||  |||| |  |||  |||||||||||||     ||| ||||
Db       1593 TGCTGCTCCTCTTTTTCATGGAGCTCATTCAGTGAGGAGTCCAGCAGCCAGAAAGGGGAG 1652

Qy         418 GATACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGAT 477
               |||||||  || || ||  | | ||||||| ||||| ||||   ||  |||  | ||  |||
Db       1653 GATACAGGCACCTGTCAGGGCCTGCCAGACAGTGAGTCCTCTTTCACATATACACTAGAT 1712

Qy         478 GAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTC 537
               |||||||||||| |||||  ||| |||| || ||||||||||   || ||  | ||||||||
Db       1713 GAAAAGGTGGCCGAGTTAGTGGAGTTCCTGCTCCTCAAATACGAAGCAGAGGAGCCTGTA 1772

Qy         538 ACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATC 597
               ||| ||||||||||||||| || |||||||   ||||| || || || |||||   ||||
Db       1773 ACAGAGGCAGAGATGCTGATGATTGTCATC---AAGTACAAAGATTACTTTCCTGTGATA 1829

Qy         598 TTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGAC 657
               ||  || ||||| |||||||||  |||||  ||||||||  |||||||||  ||  ||| |
Db       1830 CTCAAGAGAGCCCGTGAGTTCATGGAGCTTCTTTTTGGCCTTGCCCTGATAGAAGTGGGC 1889

Qy         658 CCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTG 717
               ||  |   |||||| || |  | |||| | |||||  ||| |||||||  |||||||  ||
Db       1890 CCTG---ACCACTTCTGTGTGTTTGCAAACACAGTAGGCCTCACCGATGAGGGTAG---- 1942

Qy         718 ATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTC 777
               ||||||  |||||||||||| |||||  | |||||||||  |||||||  |||  ||||||||
Db       1943 --TGATGATGAGGGCATGCCCGAGAACAGCCTCCTGATTATTATTCTGAGTGTGATCTTC 2000

Qy         778 ATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTG 837
               |||||||||| |||||  |  |||||||||||||||||||||  ||| |||| |||||||
Db       2001 ATAAAGGGCAACTGTGCCTCTGAGGAGGTCATCTGGGAAGTGCTGAATGCAGTAGGGGTA 2060

Qy         838 TGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGG 897
               | ||||||||||||||||  | ||||||| || ||  |||| |||||||  |  ||||
Db       2061 TATGCTGGGAGGGAGCACTTCGTCTATGGGGAGCCTAGGGAGCTCCTCACTAAAGTTTGG 2120

Qy         898 GTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAA 957
               ||||||| || | |||||||||||| |||||||||||| |||||  |||||||  ||||||
Db       2121 GTGCAGGGACATTACCTGGAGTATCGGGAGGTGCCCCACAGTTCTCCTCCATATTATGAA 2180

Qy         958 TTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGA--GAAGTCTTAGAGTTTTTA 1015
               ||  |||||||||||||||||||||||||    || |||||   ||||  ||||||||||
Db       2181 TTCCTGTGGGGTCCAAGAGCCCATTCAGAAAGCATCAAGAAGAAAGTACTAGAGTTTTTA 2240

Qy         1016 TCCAAGCTATCCAGTATCATCCCTAG 1041
                ||||||||   ||  |   | |||||
Db       2241 GCCAAGCTGAACAACACTGTTCCTAG 2266
```

```
RESULT 13
AF239802
LOCUS       AF239802                2887 bp    DNA     linear   PRI 02-MAR-2001
DEFINITION  Homo sapiens hepatocellular cancer antigen 587 (HCA587) gene,
            complete cds.
ACCESSION   AF239802
VERSION     AF239802.1  GI:13183142
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2887)
  AUTHORS   Qu,W., Wang,Y., Han,K.J. and Chen,W.F.
  TITLE     A new MAGE family gene identified in liver cancer
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2887)
  AUTHORS   Qu,W., Wang,Y., Han,K.J. and Chen,W.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2000) Immunology, Beijing Medical University,
            38#, Xueyuan Road, Beijing 100083, P.R. China
FEATURES             Location/Qualifiers
     source          1. .2887
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /note="found in a hepatoma patient"
     gene            1. .2887
                     /gene="HCA587"
     mRNA            join(1. .169,819. .912,1221. .2887)
                     /gene="HCA587"
     CDS             1287. .2408
                     /gene="HCA587"
                     /note="member of MAGE family"
                     /codon_start=1
                     /product="hepatocellular cancer antigen 587"
                     /protein_id="AAK15073.1"
                     /db_xref="GI:13183143"
                     /translation="MPPVPGVPFRNVDNDSPTSVELEDWVDAQHPTDEEEEEASSASS
                     TLYLVFSPSSSFSTSSSLILGGPEEEEVPSGVIPNLTESIPSSPPQGPPQGPSQSPLSS
                     CCSSFSWSSFSEESSSQKGEDTGTCQGLPDSESSFTYTLDEKVAELVEFLLLKYEAEE
                     PVTEAEMLMIVIKYKDYFPVILKRAREFMELLFGLALIEVGPDHFCVFANTVGLTDEG
                     SDDEGMPENSLLIIILSVIFIKGNCASEEVIWEVLNAVGVYAGREHFVYGEPRELLTK
                     VWVQGHYLEYREVPHSSPPYYEFLWGPRAHSESIKKKVLEFLAKLNNTVPSSFPSWYK
                     DALKDVEERVQATIDTADDATVMASESLSVMSSNVSFSE"
BASE COUNT      661 a    708 c    794 g    724 t
ORIGIN
```


```
  Query Match            46.7%;  Score 486.4;  DB 9;  Length 2887;
  Best Local Similarity  73.0%;  Pred. No. 8.7e-133;
  Matches  764;  Conservative    0;  Mismatches  211;  Indels   71;  Gaps    8;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              ||||||| | |||||  | |||||  || || | ||||  | ||||  || || | |  |
Db       1287 ATGCCTCCCGTTCCAGGCGTTCCATTCCGCAACGTTGACAACGACTCCCCGACCTCAGTT 1346

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              | |  ||| |||| ||||||||||||| || ||| |||||||| ||||||| ||||||
Db       1347 GAGTTAGAAGACTGGGTAGATGCACAGCATCCCACAGATGAGGAAGAGGAGGAAGCCTCC 1406

Qy         121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
               |||    |||||||| ||||  ||||  ||||  || |||| |||| ||
Db       1407 TCC---GCCTCTTCCACTTTGTACTTAGTATTTTCCCCCTCTTCTTTCT---------- 1452

Qy         181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
                         || | ||||| |||| ||||| |||| ||| || || |||||
Db       1453 ----------------CCACATCCTCTTCTCTGATTCTTGGTGGTCCTGAGGAGGAGGAG 1496
```



RESULT 14
AR243322
LOCUS AR243322 2940 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 20 from patent US 6475783.
ACCESSION AR243322
VERSION AR243322.1 GI:27290518
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 2940)
AUTHORS Lucas,S., De Smet,C. and Boon-Falleur,T.
TITLE Isolated nucleic acid molecule coding for tumor rejection antigen precursors MAGE-C1 and MAGE-C2 and uses thereof
JOURNAL Patent: US 6475783-A 20 05-NOV-2002;
FEATURES Location/Qualifiers



source 1. .2940
/organism="unknown"
BASE COUNT 698 a 710 c 801 g 731 t
ORIGIN

Query Match 46.7%; Score 486.4; DB 6; Length 2940;
Best Local Similarity 73.0%; Pred. No. 8.7e-133;
Matches 764; Conservative 0; Mismatches 211; Indels 71; Gaps 8;

```
Qy        958 TTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGA--GAAGTCTTAGAGTTTTTA 1015
              ||  |||||||||||||||||||||||||    || |||||    ||||  ||||||||||| 
Db       2181 TTCCTGTGGGGTCCAAGAGCCCATTCAGAAAGCATCAAGAAGAAAGTACTAGAGTTTTTA 2240

Qy       1016 TCCAAGCTATCCAGTATCATCCCTAG 1041
              |||||||   ||  |   | |||||
Db       2241 GCCAAGCTGAACAACACTGTTCCTAG 2266


RESULT 15
BD106884
LOCUS       BD106884                2940 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Isolated nucleic acid molecule coding for tumor rejection antigen
            precursors MAGE-C1 and MAGE-C2 and uses thereof.
ACCESSION   BD106884
VERSION     BD106884.1  GI:23201702
KEYWORDS    JP 2002503096-A/17.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 2940)
  AUTHORS   Lucas,S., Smet,C.D. and Falleur,T.B.
  TITLE     Isolated nucleic acid molecule coding for tumor rejection antigen
            precursors MAGE-C1 and MAGE-C2 and uses thereof
  JOURNAL   Patent: JP 2002503096-A 17 29-JAN-2002;
            LUDWIG INSTITUTE FOR CANCER RESEARCH
COMMENT     PN   JP 2002503096-A/17
            PD   29-JAN-2002
            PF   24-APR-1998 JP 1998547266
            PR   25-APR-1997 US    08/845528
            PI   SOPHIE LUCAS,CHARLES DE SMET,THIERRY BOON FALLEUR PC
            C07H21/04,A61K38/00,A61K39/00,A61K39/12,G01N33/574,C07K5/00, PC
            C07K7/00,
            PC   C07K16/00,C07K17/00
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers.
FEATURES                Location/Qualifiers
     source             1. .2940
                        /organism="synthetic construct"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:32630"
BASE COUNT      698 a    710 c    801 g    731 t
ORIGIN

  Query Match             46.7%;  Score 486.4;  DB 6;  Length 2940;
  Best Local Similarity   73.0%;  Pred. No. 8.7e-133;
  Matches  764;  Conservative    0;  Mismatches  211;  Indels   71;  Gaps    8;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              |||||||  |  |||||   |  |||||  ||  ||  |  ||||   |  ||||  ||  ||  |  |    |
Db       1287 ATGCCTCCCGTTCCAGGCGTTCCATTCCGCAACGTTGACAACGACTCCCCGACCTCAGTT 1346

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              |  |  |||  ||||  ||||||||||||||  ||  |||  |||||||||  ||||||||  ||||||
Db       1347 GAGTTAGAAGACTGGGTAGATGCACAGCATCCCACAGATGAGGAAGAGGAGGAAGCCTCC 1406

Qy        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
              |||    ||||||||  ||||  ||||  ||||  ||  ||||  ||||  ||
Db       1407 TCC---GCCTCTTCCACTTTGTACTTAGTATTTTCCCCCTCTTCTTTCT---------- 1452

Qy        181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
                          ||  |  |||||  ||||  |||||  ||||  |||  ||  ||  ||||||
Db       1453 ----------------CCACATCCTCTTCTCTGATTCTTGGTGGTCCTGAGGAGGAGGAG 1496

Qy        241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300
              ||||  |||  ||  |||  |||  ||||  ||  |||                      |
Db       1497 GTGCCCTCTGGTGTGATACCAAATCTTACCGAGA----------------------GC 1532

Qy        301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360
                |||||  |   ||||  ||||||||||||  |||  |||||  ||||||||||||   ||||
Db       1533 ATTCCCAGTAGTCCTCCACAGGGTCCTCCACAGGGTCCTTCCCAGAGTCCTCTGAGCTCC 1592

Qy        361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAG 417
              ||||  |||  ||  |||  ||||  |  |||   ||||||||||||||     |||  ||||
Db       1593 TGCTGCTCCTCTTTTTCATGGAGCTCATTCAGTGAGGAGTCCAGCAGCCAGAAAGGGGAG 1652

Qy        418 GATACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGAT 477
              |||||||  ||  ||  ||  |  |  |||||||  ||||  ||||    ||  |||  |  ||  |||
Db       1653 GATACAGGCACCTGTCAGGGCCTGCCAGACAGTGAGTCCTCTTTCACATATACACTAGAT 1712

Qy        478 GAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTC 537
              ||||||||||  ||||  |||  ||||  ||  ||||||||||   ||  ||  |  ||||||||
Db       1713 GAAAAGGTGGCCGAGTTAGTGGAGTTCCTGCTCCTCAAATACGAAGCAGAGGAGCCTGTA 1772

Qy        538 ACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATC 597
              |||  |||||||||||||||  ||  |||||||    |||||  ||  ||  ||  |||||  ||||
Db       1773 ACAGAGGCAGAGATGCTGATGATTGTCATC---AAGTACAAAGATTACTTTCCTGTGATA 1829

Qy        598 TTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGAC 657
              ||  ||  |||||  ||||||||  ||||  ||||||||  ||||||||   ||  |||  |
Db       1830 CTCAAGAGAGCCCGTGAGTTCATGGAGCTTCTTTTTGGCCTTGCCCTGATAGAAGTGGGC 1889

Qy        658 CCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTG 717
              ||  |   ||||||  ||  |  |  ||||  |  |||||  |||  |||||||  |||||||  ||
Db       1890 CCTG---ACCACTTCTGTGTGTTTGCAAACACAGTAGGCCTCACCGATGAGGGTAG---- 1942

Qy        718 ATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTC 777
                ||||||  ||||||||||  |||||  |  |||||||||  |||||||  |||  |||||||||
Db       1943 --TGATGATGAGGGCATGCCCGAGAACAGCCTCCTGATTATTATTCTGAGTGTGATCTTC 2000

Qy        778 ATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTG 837
              |||||||||  |||||  |  |  ||||||||||||||||||||||||  |||  ||||  |||||||
Db       2001 ATAAAGGGCAACTGTGCCTCTGAGGAGGTCATCTGGGAAGTGCTGAATGCAGTAGGGGTA 2060

Qy        838 TGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGG 897
              |  ||||||||||||||||  |  |||||||  ||  ||  ||||  |||||||  |  ||||
Db       2061 TATGCTGGGAGGGAGCACTTCGTCTATGGGGAGCCTAGGGAGCTCCTCACTAAAGTTTGG 2120

Qy        898 GTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAA 957
              ||||||  ||  |  ||||||||||  ||||||||||  |||||  |||||||  |||||||
Db       2121 GTGCAGGGACATTACCTGGAGTATCGGGAGGTGCCCCACAGTTCTCCTCCATATTATGAA 2180

Qy        958 TTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGA--GAAGTCTTAGAGTTTTTA 1015
              ||  |||||||||||||||||||||||||    || |||||    ||||  ||||||||||| 
Db       2181 TTCCTGTGGGGTCCAAGAGCCCATTCAGAAAGCATCAAGAAGAAAGTACTAGAGTTTTTA 2240

Qy       1016 TCCAAGCTATCCAGTATCATCCCTAG 1041
              |||||||   ||  |   | |||||
Db       2241 GCCAAGCTGAACAACACTGTTCCTAG 2266


Search completed: February 19, 2004, 19:14:08
Job time : 4193 secs
```

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 18:01:37 ; Search time 2697 Seconds
(without alignments)
9381.149 Million cell updates/sec

Title: US-10-085-108-21
Perfect score: 1041
Sequence: 1 ATGCCTCTCTTTCCAAACCT.....CTATCCAGTATCATCCCTAG 1041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	399.4	38.4	665	29	AG050338	AG050338 Pan trogl
	2	374.2	35.9	671	12	BM792325	BM792325 K-EST0072
c	3	340.8	32.7	1200	13	BX335540	BX335540 BX335540
	4	339.8	32.6	1201	13	BX335541	BX335541 BX335541
c	5	328.6	31.6	641	29	BZ608680	BZ608680 WHAAE13TF
	6	317.8	30.5	806	10	BE729944	BE729944 601562418
	7	315.4	30.3	1062	12	BM547686	BM547686 AGENCOURT
	8	306.2	29.4	861	13	BU188257	BU188257 AGENCOURT
	9	304.2	29.2	877	13	BQ434005	BQ434005 AGENCOURT
	10	302.6	29.1	937	10	BE895761	BE895761 601432579
	11	272.8	26.2	1050	10	BG024106	BG024106 602303163
	12	271.6	26.1	704	12	BI090658	BI090658 602853894
	13	271	26.0	920	13	BQ226610	BQ226610 AGENCOURT
	14	268.6	25.8	1201	13	BX365439	BX365439 BX365439
	15	265.8	25.5	820	10	BG481340	BG481340 602528684
	16	265.6	25.5	881	13	BU171902	BU171902 AGENCOURT
	17	258.6	24.8	878	10	BG397171	BG397171 602434130
	18	258.2	24.8	962	10	BG116297	BG116297 602318570
	19	256.4	24.6	776	10	BG476912	BG476912 602524881
	20	253.6	24.4	888	13	BU175305	BU175305 AGENCOURT
	21	250.4	24.1	683	10	BF979409	BF979409 602288231
	22	248.8	23.9	1973	11	AK016614	AK016614 Mus muscu
	23	248.4	23.9	886	10	BG476078	BG476078 602520963
	24	246.8	23.7	728	10	BF026372	BF026372 601671519
	25	246	23.6	497	13	BX282303	BX282303 BX282303
	26	246	23.6	1054	13	BX360398	BX360398 BX360398
	27	239	23.0	1141	12	BM473867	BM473867 AGENCOURT
	28	238	22.9	784	10	BF793468	BF793468 602254951
	29	236.4	22.7	583	10	BE392039	BE392039 601307656
	30	236.4	22.7	1066	12	BM549169	BM549169 AGENCOURT
	31	235.6	22.6	876	10	BF792356	BF792356 602252896
	32	234.6	22.5	1201	13	BX377735	BX377735 BX377735
	33	234	22.5	686	10	BG575715	BG575715 602598794
c	34	232.6	22.3	459	28	AQ017378	AQ017378 CIT-HSP-2
c	35	229	22.0	901	13	BX329175	BX329175 BX329175
	36	228.8	22.0	933	13	BX348885	BX348885 BX348885
	37	227.8	21.9	1201	9	AL530231	AL530231 AL530231
	38	225.6	21.7	1049	14	CA488474	CA488474 AGENCOURT
	39	225.6	21.7	1201	9	AL530591	AL530591 AL530591
	40	225	21.6	2163	11	AK046641	AK046641 Mus muscu
	41	224.6	21.6	1061	12	BM450221	BM450221 AGENCOURT
	42	223.6	21.5	946	10	BF983819	BF983819 602304644
	43	222.8	21.4	1201	13	BX417414	BX417414 BX417414
	44	221.6	21.3	612	10	BE890851	BE890851 601431304
	45	220	21.1	769	12	BG765526	BG765526 602739439

ALIGNMENTS

RESULT 1
AG050338/c
LOCUS AG050338 665 bp DNA linear GSS 02-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-030L19.R, genomic survey sequence.
ACCESSION AG050338
VERSION AG050338.1 GI:16587230
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 665)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of

```
           clone tracking errors.
            PRIMERS
                Sequencing: M13Rev
            LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .665
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-030L19.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      170 a    169 c    152 g    172 t      2 others
ORIGIN
```

Query Match 38.4%; Score 399.4; DB 29; Length 665;
Best Local Similarity 82.0%; Pred. No. 2.7e-85;
Matches 473; Conservative 0; Mismatches 101; Indels 3; Gaps 1;

```
Qy        300 TCCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTC 359
              || |||| |||   |||||  |  |||||  ||   |||| |||||| |||||    |||
Db        634 TCTTCCCCAGAGTGCTCCCAGGATTCCTCTTCATGATCCTACCCAGAGACCTCTGAGCTC 575

Qy        360 CTGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGA 416
              ||||||  | | |   ||||||   || ||||||||||||||||||||     |||  |||
Db        574 CTGCTCTTTCTCCTCTTTGTGAGGCCCATTGGATGAGGAGTCCAGCAGGCAAAAAGTGGA 515

Qy        417 GGATACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGA 476
              ||||  |||||||  ||||||||||||  ||||  | ||||||||||  ||  |||||  |||||
Db        514 GGATGCAGCTACATGGCATGCCTTGTCAGACAATGAATCCTTGCTCACATATGCACTGGA 455

Qy        477 TGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGT 536
              |||||||||| ||||||||||||||||||||||||||||||||||| ||| ||||||||
Db        454 TGAAAAGGTGCCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAATAAAGGAGCCTGT 395

Qy        537 CACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGAT 596
              |||||||| |||||||||||| ||||||||||||||||||||  |||||  ||||
Db        394 CACAAAGGCTGAGATGCTGACGATTGTCATCAAGAAGTATAAGGACTGCTTTCCTGTGAT 335

Qy        597 CTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGA 656
              ||||  |||||||  |||||||||||  ||  |||||  ||||  ||||||||||||||  |||||
Db        334 CTTCAGGAAAGCTCATGAGTTCATGGAACTAATGTTTGCCATTGCCCTGACTGAGATGGA 275

Qy        657 CCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCT 716
              | ||  |||||||  ||||||  ||||||||  |||||  ||||  ||  |  |||||  ||
Db        274 CACCAACAACCATTCCTATGTCTTTGAAAACACAGTAGAACTCTGCGACCAGGGGAGAAA 215

Qy        717 GATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTT 776
               | |||||  ||||||  |||  ||||||||    |||||  ||||||||||  |||  ||  ||||
Db        214 AAGTGATGGCCAGGGAATGTCCAAGAACCACCTCCTAATTCTTATTCTGAGTGTGGTCTT 155

Qy        777 CATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGT 836
              |||||||||||||||||  |  |  |||  |||||||  |||||||||  |||  |  |  ||||||||
Db        154 CATAAAGGGCAGCTGTGCCTCTGAGAAGGTCATTTGGGAAGTGCTGAATTCGATAGGGGT 95

Qy        837 GTGTGCTGGGAGGGAGCACTTTATATATGGGGATCCC 873
              ||  ||||||||||||||  |||  ||  ||||||  |  ||
Db         94 GTATGCTGGGAGGGAGCCCTTCATTTATGGGAAGACC 58
```

```
RESULT 2
BM792325
LOCUS       BM792325                671 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  K-EST0072524 S22SNU16n1 Homo sapiens cDNA clone S22SNU16n1-13-C12
            5', mRNA sequence.
ACCESSION   BM792325
VERSION     BM792325.1  GI:19140557
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 671)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 13  row: C  column: 12
            High quality sequence stop: 671.
FEATURES             Location/Qualifiers
     source          1. .671
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="S22SNU16n1-13-C12"
                     /sex="F"
                     /tissue_type="Ascites"
                     /cell_type="Lymphoblast-like"
                     /cell_line="SNU-16"
                     /lab_host="DH10B"
                     /clone_lib="S22SNU16n1"
                     /note="Organ: Stomach; Vector: pT7T3-Pac; Site_1: EcoRI;
                     Site_2: NotI; The S22SNU16 library was contributed by the
                     Soares laboratory and it was constructed as described by
                     Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome
                     Research 6(9): 791-806. RNA was prepared from harvested
                     cells of SNU-16 culture. SNU-16 cell was obtained from
                     Korean Cell Line Bank (KCLB). SNU-16 was established from
                     ascitic fluids of Korean patients by Park J.G. et al.
                     (1990), Cancer Res 50: 2773-2780."
BASE COUNT      168 a    155 c    175 g    173 t
ORIGIN
```

Query Match 35.9%; Score 374.2; DB 12; Length 671;
Best Local Similarity 82.3%; Pred. No. 3.1e-79;
Matches 455; Conservative 0; Mismatches 93; Indels 5; Gaps 2;

```
Qy        490 GAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAG 549
              |||||||  || |||||||||||||||||||||||    ||  ||||| ||||||||||||||
Db          4 GAGTTGGCGCGGTTTCTTCTCCTCAAATATCAAGTGAAGCAGCCTATCACAAAGGCAGAG 63

Qy        550 ATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCC 609
              ||||||||||  ||||||||   | |||  | ||  |||  |||||   ||||||||  ||||||||
Db         64 ATGCTGACGAATGTCATCAGCAGGTACACGGGCTACTTTCCTGTGATCTTCAGGAAAGCC 123

Qy        610 CATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCAC 669
              | ||||||||||||| ||  ||||||||||  ||||||   ||  |||||||  |  |  ||
Db        124 CGTGAGTTCATAGAGATACTTTTTGGCATTTCCCTGAGAGAAGTGGACCCTG---ATGAC 180

Qy        670 TCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAG 729
              ||||||  ||||||  | ||||||||||||||||||  ||||||   | ||||  ||||||  |||
Db        181 TCCTATGTCTTTGTAAACACATTAGACCTCACCTCTGAGGGGTGTCTGAGTGATGAGCAG 240

Qy        730 GGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGC 789
              ||||||  ||  ||||||  | ||||||||||||||||  |||||  |||||||||||||  |
Db        241 GGCATGTCCCAGAACCGCCTCCTGATTCTTATTCTGAGTATCATCTTCATAAAGGGCACC 300

Qy        790 TGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGG 849
              | ||  | |  ||||||||||||||||||  |||  ||||||  |||||||||  ||||||||||
Db        301 TATGCCTCTGAGGAGGTCATCTGGGATGTGCTGAGTGGAATAGGGGTGCGTGCTGGGAGG 360

Qy        850 GAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTGCAGAGAAAG 909
              |||||||||    |  ||||||  ||||  ||||  |||||||  |  ||||||||||  |  ||
Db        361 GAGCACTTTGCCTTTGGGGAGCCCAGGGAGCTCCTCACTAAAGTTTGGGTGCAGGAACAT 420
```

```
Qy        910 TACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAATTTTTGTGGGGT 969
              ||||| |||||||||||||||||||||||  | ||||||| ||||| |||||  |||||||
Db        421 TACCTAGAGTACCGGGAGGTGCCCAACTCTTCTCCTCCTCGTTACGAATTCCTGTGGGGT 480

Qy        970 CCAAGAGCCCATTCAGAGGCCAGCAAGAG--AAGTCTTAGAGTTTTTATCCAAGCTATCC 1027
              |||||||| ||||||||  | ||  |||||  ||||  ||||||||||  ||| ||||
Db        481 CCAAGAGCTCATTCAGAAGTCATTAAGAGGAAAGTAGTAGAGTTTTTGGCCATGCTAAAG 540

Qy       1028 AGTATCATCCCTA 1040
              | || | ||||||
Db        541 AATACCGTCCCTA 553


RESULT 3
BX335540/c
LOCUS       BX335540               1200 bp    mRNA    linear   EST 02-MAY-2003
DEFINITION  BX335540 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI016YG21 3-PRIME, mRNA sequence.
ACCESSION   BX335540
VERSION     BX335540.1  GI:30335452
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1200)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 10428.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI016AD11NP1&cluster=10428.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI016AD11NP1.
FEATURES             Location/Qualifiers
     source          1. .1200
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI016YG21"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      262 a    298 c    281 g    285 t     74 others
ORIGIN

  Query Match           32.7%;  Score 340.8;  DB 13;  Length 1200;
  Best Local Similarity  71.9%;  Pred. No. 4e-71;
  Matches  460;  Conservative    0;  Mismatches  177;  Indels    3;  Gaps    1;

Qy        368 CCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAGAGGAGGA---TACAG 424
              | || |||   |    | ||  |||||||  ||||||||  || |||||||   | ||
Db        833 CTCCCCTTCCATTAGATCAATCTGATGAGGGCTCCAGCAGCCAAAAGGAGGAGAGTCCAA 774

Qy        425 CTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGG 484
              ||    || | | ||||||| ||||| || || |||||    ||    | ||||||||||
Db        773 GCACCCTACAGGTCCTGCCAGACAGTGAGTCTTTACCCAGAAGTGAGATAGATGAAAAGG 714

Qy        485 TGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGG 544
              || |||| |||||||||||||| ||| |||| |||||||  || |||||  |||||||||
Db        713 TGACTGATTTGGTGCAGTTTCTGCTCTTCAAGTATCAAATGAAGGAGCCGATCACAAAGG 654

Qy        545 CAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGA 604
              |||| || |||  || |||||| || || ||| | ||| | || ||  || | ||  |
Db        653 CAGAAATACTGGAGAGTGTCATAAAAAATTATGAAGACCACTTCCCTTTGTTGTTTAGTG 594

Qy        605 AAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACA 664
              |||||   |||| |||   ||| | ||||||||||   |  |  ||  |||| |||
Db        593 AAGCCTCCGAGTGCATGCTGCTGGTCTTTGGCATTGATGTAAAGGAAGTGGATCCCACTG 534

Qy        665 ACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATG 724
              |||||||| | || |||    | |  | | |||||||||||| || |  |||| |||||
Db        533 GCCACTCCTTTGTCCTTGTCACCTCCCTGGGCCTCACCTATGATGGGATGCTGAGTGATG 474

Qy        725 ACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGG 784
              |||| ||||||||||||    | | || || ||||| || || || |||||||||| |||
Db        473 TCCAGAGCATGCCCAAGACTGGCATTCTCATACTTATCCTAAGCATAATCTTCATAGAGG 414

Qy        785 GCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTG 844
              ||  |||   ||| |||||||||||||||||||    ||| |   || ||| ||| || ||
Db        413 GCTACTGCACCCCTGAGGAGGTCATCTGGGAAGCACTGAATATGATGGGGCTGTATGATG 354

Qy        845 GGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTGCAGA 904
              ||| |||||||  | || |||||||| |||| ||||||||||    | ||||||||||
Db        353 GGATGGAGCACCTCATTTATGGGGAGCCCAGGAAGCTGCTCACCCAAGATTGGGTGCAGG 294

Qy        905 GAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAATTTTTGT 964
              ||| |||||||||||||| |||||||  |||||  |||| |||| |||||  ||| |||
Db        293 AAAACTACCTGGAGTACCGGCAGGTGCCTGGCAGTGATCCTGCACGGTATGAGTTTCTGT 234

Qy        965 GGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAGAAGTCT 1004
              |||||||||| || ||| | ||    || ||||  |||||
Db        233 GGGGTCCAAGGGCTCATGCTGAAATTAGGAAGATGAGTCT 194


RESULT 4
BX335541
LOCUS       BX335541               1201 bp    mRNA    linear   EST 02-MAY-2003
DEFINITION  BX335541 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI016YG21 5-PRIME, mRNA sequence.
ACCESSION   BX335541
VERSION     BX335541.1  GI:30337426
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 10428.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI016AD11QP1&cluster=10428.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI016AD11QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI016YG21"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
```

```
BASE COUNT      321 a    256 c    289 g    282 t     53 others
ORIGIN

  Query Match             32.6%;  Score 339.8;  DB 13;  Length 1201;
  Best Local Similarity   72.7%;  Pred. No. 6.9e-71;
  Matches  453;  Conservative    0;  Mismatches  167;  Indels    3;  Gaps    1;

Qy        385 CGATTGGATGAGGAGTCCAGCAGTGAAGAGGAGGA---TACAGCTACTTGGCATGCCTTG 441
              | ||  |||||||  ||||||||  || |||||||    | ||   ||     || | | ||
Db        152 CAATCTGATGAGGGCTCCAGCAGCCAAAAGGAGGAGAGTCCAAGCACCCTACAGGTCCTG 211

Qy        442 CCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTGGTGCAG 501
              ||||| ||||| || || |||||   ||   | ||||||||||| |||| |||||||||
Db        212 CCAGACAGTGAGTCTTTACCCAGAAGTGAGATAGATGAAAAGGTGACTGATTTGGTGCAG 271

Qy        502 TTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTGACGACT 561
              ||||| ||| |||| |||||||   || |||||  |||||||||||| || |||  || |
Db        272 TTTCTGCTCTTCAAGTATCAAATGAAGGAGCCGATCACAAAGGCAGAAATACTGGAGAGT 331

Qy        562 GTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAGTTCATA 621
              ||||| || || ||| | ||| | || ||  || | ||  |  |||||   |||| |||
Db        332 GTCATAAAAAATTATGAAGACCACTTCCCTTTGTTGTTTAGTGAAGCCTCCGAGTGCATG 391

Qy        622 GAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTATTTCTTT 681
                ||| | ||||||||||    | |  ||  |||| |||      |||||||| | || ||
Db        392 CTGCTGGTCTTTGGCATTGATGTAAAGGAAGTGGATCCCACTGGCCACTCCTTTGTCCTT 451

Qy        682 GAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATGCCCAAG 741
              |    | |  | | |||||||||||| || |  |||| |||| |||| ||||||||||
Db        452 GTCACCTCCCTGGGCCTCACCTATGATGGGATGCTGAGTGATGTCCAGAGCATGCCCAAG 511

Qy        742 AACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTCCCCGAG 801
              |    |  || || |||| || || || |||||||| ||||  |||   ||| |||
Db        512 ACTGGCATTCTCATACTTATCCTAAGCATAATCTTCATAGAGGGCTACTGCACCCCTGAG 571

Qy        802 GAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCACTTTATA 861
              |||||||||||||||    ||| |   || ||| ||| || |||| |||||||  | ||
Db        572 GAGGTCATCTGGGAAGCACTGAATATGATGGGGCTGTATGATGGGATGGAGCACCTCATT 631

Qy        862 TATGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTGCAGAGAAAGTACCTGGAGTAC 921
              |||||||| |||| |||||||||||   | ||||||||||  ||| ||||||||||||
Db        632 TATGGGGAGCCCAGGAAGCTGCTCACCCAAGATTGGGTGCAGGAAAACTACCTGGAGTAC 691

Qy        922 CGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAATTTTTGTGGGGTCCAAGAGCCCAT 981
              ||| |||||||    |||||  |||| |||| |||||  ||| ||||||||||||| || |||
Db        692 CGGCAGGTGCCTGGCAGTGATCCTGCACGGTATGAGTTTCTGTGGGGTCCAAGGGCTCAT 751

Qy        982 TCAGAGGCCAGCAAGAGAAGTCT 1004
              | |||     || ||||  |||||
Db        752 GCTGAAATTAGGAAGATGAGTCT 774


RESULT 5
BZ608680/c
LOCUS       BZ608680                641 bp    DNA     linear   GSS 08-JUN-2003
DEFINITION  WHAAE13TF Human MCF7 breast cancer cell line library (MCF7_1) Homo
            sapiens genomic clone MCF7_1-2C1, genomic survey sequence.
ACCESSION   BZ608680
VERSION     BZ608680.1  GI:31517241
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,
            Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,
            Gray,J.W. and Collins,C.
  TITLE     ESP: a sequence-based approach to the structural genomics of tumors
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Volik SV
            Colin Collins' lab
            UCSF Comprehensive Cancer Center
            UCSF Box 0808, San Francisco, CA 94143-0808, USA
            Tel: 415 502 7066
            Fax: 415 502 5665
            Email: svolik@cc.ucsf.edu
            This clone is available from Amplicon Express
            http://www.genomex.com
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .641
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="MCF7_1-2C1"
                     /sex="female"
                     /clone_lib="Human MCF7 breast cancer cell line library
                     (MCF7_1)"
                     /note="Vector: pECBAC1; Site_1: HindIII; This library was
                     constructed from MCF7 breast cancer cell line by Amplicon
                     Express (http://www.genomex.com) using their standard
                     procedure."
BASE COUNT      161 a    166 c    149 g    165 t
ORIGIN

  Query Match             31.6%;  Score 328.6;  DB 29;  Length 641;
  Best Local Similarity   81.4%;  Pred. No. 2.8e-68;
  Matches  445;  Conservative    0;  Mismatches   89;  Indels   13;  Gaps    5;

Qy        496 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 555
              ||| | || || ||||||||||||||| |||| |||||| | ||| || ||||||||||
Db        641 GTGGATTTCCTGCTCCTCAAATATCAACCAAAGGAGCCTATAACAGAGTCAGAGATGCTG 582

Qy        556 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 615
              |||| ||||||||  |||||  || |||||  ||| ||  ||||  |||  |||||||  ||||
Db        581 ACGATTGTCATCACCAAGTACAAAGACTACTTT-CCGTGATACTCGAGAAAGCCTGTGAG 523

Qy        616 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 675
              |||||||| |  |||||||| || |||||| ||  |||  |||  |   |||| | ||||
Db        522 TTCATAGAG-TTCTTTTTGGCTTTTCCCTGACAGAAGTGGGCCCTG---ACCATTTCTAT 467

Qy        676 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 735
              | |||| | |||||  |||||||| || |||||| ||      |||||||  |||| |
Db        466 GTGTTTGCAAACACAGTAGACCTCGCCGATGAGGGTAG------TGATGACACGGGCAGG 413

Qy        736 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 795
              ||  ||||||| |||||||||||||||| ||| |||||||||||| |||||||||| |
Db        412 CCTGAGAACTGCCTCCTGATTCTTATTCTGAGTGTGATCTTCATAACGGGCAGCTGTGCC 353

Qy        796 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 855
              || |||||||||||||||||||| |||  |||||||||||| ||||||||||||||||
Db        352 CCTGAGGAGGTCATCTGGGAAGTGCTGAAGGCAATAGGGGTGTATGCTGGGAGGGAGCAC 293

Qy        856 TTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTGCAGAGAAAGTACCTG 915
              ||  | |||||||| |||||  |||| |||||||  |  |||||||||   | | ||||||
Db        292 TTCGTCTATGGGGAGCCCAGGGAGCTCCTCACTAAAGTTTGGGTGCAGGAACATTACCTG 233

Qy        916 GAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAATTTTTGTGGGGTCCAAGA 975
              |||||  |||||||||||||||||  ||||||||||||||||||||  ||||||||||||||
Db        232 GAGTATTGGGAGGTGCCCAACAGTTCTCCTCCACGTTATGAATTCCTGTGGGGTCCAAGA 173

Qy        976 GCCCATTCAGAGGCCAGCAAG--AGAAGTCTTAGAGTTTTTATCCAAGCTATCCAGTATC 1033
              ||||||||||    || ||||    |||| ||||||||  ||||||||  || ||
Db        172 GCCCATTCAGAAAGCACCAAGCAGAAAGTACTAGAGTTTTTGGCCAAGCTATACAATAGT 113

Qy       1034 ATCCCTA 1040
              |||||||
Db        112 ATCCCTA 106


RESULT 6
BE729944
LOCUS       BE729944                806 bp    mRNA    linear   EST 15-SEP-2000
DEFINITION  601562418F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3831893 5',
            mRNA sequence.
```

```
ACCESSION   BE729944
VERSION     BE729944.1  GI:10143936
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 806)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM508  row: d  column: 06
            High quality sequence stop: 707.
FEATURES             Location/Qualifiers
     source          1. .806
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3831893"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_20"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      186 a    203 c    217 g    199 t      1 others
ORIGIN
```

```
  Query Match           30.5%;  Score 317.8;  DB 10;  Length 806;
  Best Local Similarity   69.2%;  Pred. No. 1.2e-65;
  Matches  580;  Conservative    0;  Mismatches  188;  Indels   70;  Gaps    8;

Qy         18 CCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGTGTGACAGAGGACTTGGT 77
              | |||||  || || | ||||  | |||| || || | |   || |  ||| |||| |||
Db          2 CGTTCCATTCCGCAACGTTGACAACGACTCCCCGACCTCAGTTGAGTTAGAAGACTGGGT 61

Qy         78 AGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCCTCCACTTCCTCTTCCTC 137
              |||||||||| || ||| |||||||| |||||||| ||||||||     |||||||| |
Db         62 AGATGCACAGCATCCCACAGATGAGGAAGAGGAGGAAGCCTCCTCC---GCCTCTTCCAC 118

Qy        138 TTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCCTCACCCTTGTCCTCACC 197
              |||  ||||  |||| || |||| |||| ||                           |
Db        119 TTTGTACTTAGTATTTTCCCCCTCTTCTTTCT---------------------------C 151

Qy        198 CTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGATATGCCTGCTGCTGGGAT 257
              |  | ||||| |||| ||||| |||| ||| || || |||||  ||||  ||| || |||
Db        152 CACATCCTCTTCTCTGATTCTTGGTGGTCCTGAGGAGGAGGAGGTGCCCTCTGGTGTGAT 211

Qy        258 GCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGTCCTCCCAAGATCTCTCC 317
               |||  |||| || |||                        |  |||||  |  ||||
Db        212 ACCAAATCTTACCGAGA----------------------GCATTCCCAGTAGTCCTCC 247

Qy        318 CCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCATCCCCTCTTTT 377
              ||||||||||| ||| ||||| |||||||||||||   ||||||||  ||| || |||
Db        248 ACAGGGTCCTCCACAGGGTCCTTCCCAGAGTCCTCTGAGCTCCTGCTGCTCCTCTTTTTC 307

Qy        378 GTGGACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAGGATACAGCTACTTGGCA 434
              ||||  |  |||   ||||||||||||||     |||  ||||||||||  || || ||
Db        308 ATGGAGCTCATTCAGTGAGGAGTCCAGCAGCCAGAAAGGGGAGGATACAGGCACCTGTCA 367

Qy        435 TGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTT 494
              | | ||||||| |||| |||| ||    ||  ||| | || ||||||||||||| |||||
Db        368 GGGCCTGCCAGACAGTGAGTCCTCTTTCACATATACACTAGATGAAAAGGTGGCCGAGTT 427

Qy        495 GGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCT 554
              ||| |||| || |||||||||||  || || | |||||||| ||| |||||||||||||
Db        428 AGTGGAGTTCCTGCTCCTCAAATACGAAGCAGAGGAGCCTGTAACAGAGGCAGAGATGCT 487

Qy        555 GACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGA 614
              || || |||||||   ||||| || || || |||||  ||||  ||  || ||||| |||
Db        488 GATGATTGTCATC---AAGTACAAAGATTACTTTCCTGTGATACTCAAGAGAGCCCGTGA 544

Qy        615 GTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTA 674
              ||||| |||||  |||||||| |||||||||  ||  ||| ||| |   |||||| ||
Db        545 GTTCATGGAGCTTCTTTTTGGCCTTGCCCTGATAGAAGTGGGCCCTG---ACCACTTCTG 601

Qy        675 TTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCAT 734
              | | |||| | ||||| ||| ||||||| |||| ||        | ||||||  |||||||
Db        602 TGTGTTTGCAAACACAGTAGGCCTCACCGATGANGG-----GTAGTGATGATGAGGGCAT 656

Qy        735 GCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGT 794
              |||| ||||| | |||||||| ||||||| ||| ||||||||||||||||
Db        657 GCCCGAGAACAGGCTCCTGATTATTATTCTGAGTGTGATCTTCATAAAGGGGCAACTGTG 716

Qy        795 CC--CCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGG 850
              ||  | || |||||||||||||||||| ||| |||    ||| | || |   ||  ||
Db        717 CCTTCTGAAGAGGTCATCTGGGAAGTGCTGAATGCCGAGGGGATATGCGGGAGGGAGG 774
```

```
RESULT 7
BM547686
LOCUS       BM547686               1062 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6507236 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5727806
            5', mRNA sequence.
ACCESSION   BM547686
VERSION     BM547686.1  GI:18781667
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1062)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12722  row: h  column: 15
            High quality sequence stop: 631.
FEATURES             Location/Qualifiers
     source          1. .1062
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5727806"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_124"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: EcoRV
                     (destroyed); Site_2: NotI; RNA source male hippocampus,
                     age 27. Library is oligo-dT primed and directionally
                     cloned (EcoRV site is destroyed upon cloning). Average
                     insert size 1.4 kb, insert size range 0.9-4 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 012."
```

```
BASE COUNT      246 a    330 c    254 g    232 t
ORIGIN

  Query Match           30.3%;  Score 315.4;  DB 12;  Length 1062;
  Best Local Similarity   84.3%;  Pred. No. 5e-65;
  Matches  451;  Conservative    0;  Mismatches   71;  Indels   13;  Gaps    8;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        525 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 584

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGG-AGGAGGAGGATGCCTC 119
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        585 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAAGGAGGAGGATGCCTC 644

Qy        120 CTCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATC 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        645 CTCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATC 704

Qy        180 CTCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGT-GTTCCAGAAGATGAGG 238
              |||||||||||||||||||||||||||||||||||||||||| |||||| |||||||||
Db        705 CTCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGGTTCCAAAAGATGAGG 764

Qy        239 ATATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGG 298
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        765 ATATGCCTGCTGCTGGGATGCCACCTCTTTCCCAGAGTCCTCCTGAGATTCCTCCCCAGG 824

Qy        299 GT-CCTCCCAAGATCTCTCCCCAGGGT-CCTCCGCAGAGTCCTCCCCAGAG-TCCTCTAG 355
              || |||||||  |||||||||||||| |||||||| | ||||||||  ||  ||||||
Db        825 GTTCCTCCCAGAATCTCTCCCCAGGGTCCCTCCGCAAAATCCTCCCCCCAGAGCCTCTAA 884

Qy        356 ACTCCTGCTCATCCCCTCTTTTGTGG----ACCCGATTGGATGAGGAGTCCAGC---AGT 408
              | ||||| |||| |||   ||| |||     |||  | | |||||||  |  |    |
Db        885 ATTCCTGGTCATTCCCCTCTTTTTGGGGACCCCCATTGGAATGAGGAACTCCACCCAGGA 944

Qy        409 GAAGAGGAGGATACAGCTACTTGGCA-TGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTA 467
              || || ||||||     |||||||||| | |||  ||  |||| ||||||||||| || |
Db        945 AAAAAGCAGGATCACCTTACTTGGCAGTCCCTGCCCCAAAAGGGAATCCTTGCCCCGGAA 1004

Qy        468 TGCCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAA 522
              | |||  ||  ||||| |||    |   ||  |  | | | |||   |  |||
Db       1005 TTCCCCCGAAAAAAAGATGGCCTCGACTTTGGTGCCACCTTTTCTCCCCTCACAA 1059


RESULT 8
BU188257
LOCUS       BU188257                 861 bp    mRNA    linear   EST 04-SEP-2002
DEFINITION  AGENCOURT_7931540 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6150482
            5', mRNA sequence.
ACCESSION   BU188257
VERSION     BU188257.1  GI:22702241
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 861)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13485  row: d  column: 03
            High quality sequence stop: 596.
FEATURES             Location/Qualifiers
     source          1. .861
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6150482"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_72"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      191 a    220 c    235 g    215 t
ORIGIN

  Query Match           29.4%;  Score 306.2;  DB 13;  Length 861;
  Best Local Similarity   70.6%;  Pred. No. 7.6e-63;
  Matches  580;  Conservative    0;  Mismatches  168;  Indels   73;  Gaps   10;

Qy          3 GCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGTGT 62
              ||||| | ||||  | ||||  || || | ||||  | |||| || || | |   ||
Db          1 GCCTCCCGTTCCAGGCGTTCCATTCCGCAACGTTGACAACGACTCCCCGACCTCAGTTGA 60

Qy         63 GACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCCTC 122
              |  ||| |||| |||||||||||||| || ||| |||||||| |||||||| |||||||
Db         61 GTTAGAAGACTGGGTAGATGCACAGCATCCCACAGATGAGGAAGAGGAGGAAGCCTCCTC 120

Qy        123 CACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCCTC 182
              |   |||||||| ||||  ||||  ||||  || |||| |||| ||
Db        121 C---GCCTCTTCCACTTTGTACTTAGTATTTTCCCCCTCTTCTTTCT------------- 164

Qy        183 ACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGATAT 242
                          ||  | ||||| |||| ||||| |||| ||| || || ||||| |
Db        165 --------------CCACATCCTCTTCTCTGATTCTTGGTGGTCCTGAGGAGGAGGAGGT 210

Qy        243 GCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGTCC 302
              |||  ||| || ||| |||  |||| || ||||   |||     | |||||
Db        211 GCCCTCTGGTGTGATACCAAATCTTACCGAGAGCATTCCCAGTAGTCCTC---------- 260

Qy        303 TCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTG 362
                          | |||||||||||| ||| |||| ||||||||||||    ||||||
Db        261 --------------CACAGGGTCCTCCACAGGGTCCTTCCCAGAGTCCTCTGAGCTCCTG 306

Qy        363 CTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAGGA 419
              ||  ||| || |||  |||| |  |||  |||||||||||||     ||| ||||||
Db        307 CTGCTCCTCTTTTTCATGGAGCTCATTCAGTGAGGAGTCCAGCAGCCAGAAAGGGGAGGA 366

Qy        420 TACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGA 479
              |||||  || || ||  | | ||||||| |||||  ||||     ||  ||| | || |||||
Db        367 TACAGGCACCTGTCAGGGCCTGCCAGACAGTGAGTCCTCTTTCACATATACACTAGATGA 426

Qy        480 AAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCAC 539
              |||||||| |||||  ||| |||| || ||||||||||   || || | ||||||| ||
Db        427 AAAGGTGGCCGAGTTAGTGGAGTTCCTGCTCCTCAAATACGAAGCAGAGGAGCCTGTAAC 486

Qy        540 AAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTT 599
              | |||||||||||||| || |||||||   |||||  || || || |||||  ||||  |
Db        487 AGAGGCAGAGATGCTGATGATTGTCATC---AAGTACAAAGATTACTTTCCTGTGATACT 543

Qy        600 CGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCC 659
              |  ||  ||||| |||||||||  |||||  ||||||||  |||||||||   ||  ||| |||
Db        544 CAAGAGAGCCCGTGAGTTCATGGAGCTTCTTTTTGGCCTTGCCCTGATAGAAGTGGGCCC 603

Qy        660 CGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGAT 719
              |   ||||||  || | | |||| | ||||| ||| |||||||  |||||||  ||
Db        604 TG---ACCACTTCTGTGTGTTTGCAAACACAGTAGGCCTCACCGATGAGGGTAG------ 654

Qy        720 TGATGACCAGGGCATGCCCAAGAACTGTCT-CCTGATTCTTATTCTCAGTATGATCTTCA 778
              ||||||  |||||||||||  ||||| | || |||||||  |||||||  ||| |||||||||
Db        655 TGATGATGAGGGCATGCCCGAGAACAGCCTCCCTGATTATTATTCTGAGTGTGATCTTCA 714

Qy        779 TAAA-GGGCAGCTGTGTCCCC--GAGGAGGTCATCTGGGAA 816
              |||| |||||  |||||  || |  |||||||||||||||| |
Db        715 TAAAGGGGCAACTGTGGCCTCTGGAGGAGGTCATCTGGGGA 755
```

```
RESULT 9
BQ434005
LOCUS       BQ434005                877 bp    mRNA    linear   EST 24-MAY-2002
DEFINITION  AGENCOURT_7790940 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6065636
            5', mRNA sequence.
ACCESSION   BQ434005
VERSION     BQ434005.1  GI:21173081
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 877)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13342  row: d  column: 21
            High quality sequence stop: 647.
FEATURES             Location/Qualifiers
     source          1. .877
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6065636"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_72"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      203 a    233 c    254 g    185 t      2 others
ORIGIN

  Query Match             29.2%;  Score 304.2;  DB 13;  Length 877;
  Best Local Similarity   70.6%;  Pred. No. 2.3e-62;
  Matches  433;  Conservative    0;  Mismatches  175;  Indels    5;  Gaps    2;

Qy        375 TTTGTGGACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAGGATACAGCTACTTG 431
              | ||||||| || ||  |||||||  ||||||||   ||||||||||    ||   || |
Db        238 TCTGTGGAGCCAATCCGATGAGGGTTCCAGCAGCAATGAAGAGGAGGGGCCAAGCACCTC 297

Qy        432 GCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGA 491
              | | |     |  |     || ||| ||  | || | || || ||||| || ||||||||
Db        298 CCCGGACCCAGCTCACCTGGAGTCCCTGTTCCGGGAAGCACTTGATGAGAAAGTGGCTGA 357

Qy        492 GTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGAT 551
              ||| || |  || || |||| |||||||||||  || ||||| ||||||||||||||| ||
Db        358 GTTAGTTCGTTTCCTGCTCCGCAAATATCAAATTAAGGAGCCGGTCACAAAGGCAGAAAT 417

Qy        552 GCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCA 611
              |||   || |||||||| || || ||| || | |||||    |||||| | ||||||
Db        418 GCTTGAGAGTGTCATCAAAAATTACAAGAACCACTTTCCTGATATCTTCAGCAAAGCCTC 477

Qy        612 TGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTC 671
              ||||| |||  || | || |||||||||    |||  ||  |||||| | |  ||||||
Db        478 TGAGTGCATGCAGGTGATCTTTGGCATTGATGTGAAGGAAGTGGACCCTGCCGGCCACTC 537

Qy        672 CTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGG 731
              |||  || |||    |  | | |||| ||||||| ||    |||  ||||||  ||| |
Db        538 CTACATCCTTGTCACCTGCCTGGGCCTCTCCTATGATGGCCTGCTGGGTGATGATCAGAG 597

Qy        732 CATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTG 791
              | |||||||| | | |||||||  |  | ||  | |||||||| ||  |||||||| |
Db        598 TACGCCCAAGACCGGCCTCCTGATAATCGTCCTGGGCATGATCTTAATGGAGGGCAGCCG 657

Qy        792 TGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGA 851
              | ||| |||||||  |||||||||| ||||||||  || ||| ||| || |||||||||
Db        658 CGCCCCGGAGGAGGCAATCTGGGAAGCGTTGAGTGTGATGGGGCTGTATGATGGGAGGGA 717

Qy        852 GCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTGCAGAGAAAGTA 911
              ||||  | | ||| || | | ||| ||||||||||    | | |||||||||   || ||
Db        718 GCACAGTGTCTATTGGAAGCTCAGGAAGCTGCTCACCCAAGAGTGGGTGCAGGAGAACTA 777

Qy        912 CCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAATTTT--TGTGGGGT 969
              ||||||||||||  ||| |||| ||||| |||| || ||  ||  ||  |||||||
Db        778 CCTGGAGTACCGCCAGGCGCCCGGCAGTGATCCTGTGCGCTACCAANNTTCCGGTGGGGT 837

Qy        970 CCAAGAGCCCATT 982
              ||||| | || ||
Db        838 CCAAGGGGCCCTT 850


RESULT 10
BE895761
LOCUS       BE895761                937 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601432579F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3917778 5',
            mRNA sequence.
ACCESSION   BE895761
VERSION     BE895761.1  GI:10359482
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 937)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9744  row: n  column: 19
            High quality sequence stop: 619.
FEATURES             Location/Qualifiers
     source          1. .937
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3917778"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_72"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      218 a    189 c    269 g    261 t
ORIGIN

  Query Match             29.1%;  Score 302.6;  DB 10;  Length 937;
  Best Local Similarity   81.2%;  Pred. No. 5.7e-62;
  Matches  389;  Conservative    0;  Mismatches   84;  Indels    6;  Gaps    3;

Qy        564 CATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAGTTCATAGA 623
              |||||  | ||| | || ||| |||||  |||||||| |||||||| ||||||||||||
Db          1 CATCAGCAGGTACACGGGCTACTTTCCTGTGATCTTCAGGAAAGCCCGTGAGTTCATAGA 60

Qy        624 GCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTATTTCTTTGA 683
```

```
Db         61 GATACTTTTTGGCATTTCCCTGAGAGAAGTGGACCCTG---ATGACTCCTATGTCTTTGT 117

Qy        684 AGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATGCCCAAGAA 743

Db        118 AAACACATTAGACCTCACCTCTGAGGGGTGTCTGAGTGATGAGCAGGGCATGTCCCAGAA 177

Qy        744 CTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGA 803

Db        178 CCGCCTCCTGATTCTTATTCTGAGTATCATCTTCATAAAGGGCACCTATGCCTCTGAGGA 237

Qy        804 GGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCACTTTATATA 863

Db        238 GGTCATCTGGGATGTGCTGAGTGGAATAGGGGTGCGTGCTGGGAGGGAGCACTTTGCCTT 297

Qy        864 TGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTGCAGAGAAAGTACCTGGAGTACCG 923

Db        298 TGGGGAGCCCAG-GAGCTCCTCACTAAAGTTTGGGTGCAGGAACATTACCTAGAGTACCG 356

Qy        924 GGAGGTGCCCAACAGTGCTCCTCCACGTTATGAATTTTTGTGGGGTCCAAGAGCCCATTC 983

Db        357 GGAAGTGCCCAACTCTTCTCCTCCTCGTTACGAATTCCTGTGGGGTCCAAGAGCTCATTC 416

Qy        984 AGAGGCCAGCAAGAG--AAGTCTTAGAGTTTTTATCCAAGCTATCCAGTATCATCCCTA 1040

Db        417 AGAAGTCATTAAGAGGAAAGTAGTAGAGTTTTTGGCCATGCTAAAGAATACCGTCCCTA 475


RESULT 11
BG024106
LOCUS       BG024106                1050 bp    mRNA    linear   EST 24-JAN-2001
DEFINITION  602303163F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4394703 5',
            mRNA sequence.
ACCESSION   BG024106
VERSION     BG024106.1  GI:12409339
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1050)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10090  row: n  column: 16
            High quality sequence stop: 700.
FEATURES             Location/Qualifiers
     source          1. .1050
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4394703"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_88"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      276 a    246 c    332 g    196 t
ORIGIN

  Query Match             26.2%;  Score 272.8;  DB 10;  Length 1050;
  Best Local Similarity   69.1%;  Pred. No. 8.7e-55;
  Matches  387;  Conservative    0;  Mismatches  172;  Indels    1;  Gaps    1;

Qy        437 CCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTGG 496

Db         38 CCTCGCCTGACGCAGAGTCCTTGTTCCGAGAAGCACTCAGTAACAAGGTGGATGAGTTGG 97

Qy        497 TGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTGA 556

Db         98 CTCATTTTCTGCTCCGCAAGTATCGAGCCAAGGAGCTGGTCACAAAGGCAGAAATGCTGG 157

Qy        557 CGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAGT 616

Db        158 AGAGAGTCATCAAAAATTACAAGCGCTGCTTTCCTGTGATCTTCGGCAAAGCCTCCGAGT 217

Qy        617 TCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTATT 676

Db        218 CCCTGAAGATGATCTTTGGCATTGACGTGAAGGAAGTGGACCCCACCAGCAACACCTACA 277

Qy        677 TCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATGC 736

Db        278 CCCTTGTCACCTGCCTGGGCCTTTCCTATGATGGCCTGCTGGGTAATAATCAGATCTTTC 337

Qy        737 CCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTCC 796

Db        338 CCAAGACAGGCCTTCTGATAATCGTCCTGGGCACAATTGCAATGGAGGGCGACAGCGCCT 397

Qy        797 CCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCACT 856

Db        398 CTGAGGAGGAAATCTGGGAGGAGCTGGGTGTGATGGGGGTGTATGATGGGAGGGAGCACA 457

Qy        857 TTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTGCAGAGAAAGTACCTGG 916

Db        458 CTGTCTATGGGGAGCCCAGGAAACTGCTCACCCAAGATTGGGTGCAGGAAAACTACCTGG 517

Qy        917 AGTACCGGGAGGTGCCCAACAGT-GCTCCTCCACGTTATGAATTTTTGTGGGGTCCAAGA 975

Db        518 AGTACCGGCAGGTACCCGGCAGTAAATCCTGCGCGCTATGAGTTCCTGTGGGGTCCAAGG 577

Qy        976 GCCCATTCAGAGGCCAGCAA 995

Db        578 GCTCTGGCTGAAACCAGCTA 597


RESULT 12
BI090658
LOCUS       BI090658                704 bp    mRNA    linear   EST 20-JUN-2001
DEFINITION  602853894F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4995238 5',
            mRNA sequence.
ACCESSION   BI090658
VERSION     BI090658.1  GI:14508988
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 704)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11018  row: l  column: 23
            High quality sequence stop: 703.
FEATURES             Location/Qualifiers
     source          1. .704
```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4995238"
/cell_line="MGC36"
/lab_host="DH10B"
/clone_lib="NIH_MGC_10"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5 kb. Library prepared by Life Technologies."

BASE COUNT 158 a 174 c 214 g 157 t 1 others
ORIGIN

Query Match 26.1%; Score 271.6; DB 12; Length 704;
Best Local Similarity 67.8%; Pred. No. 1.5e-54;
Matches 379; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

```
Qy        437 CCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTGG 496
              |||| || ||    || |||  | |     || ||   | ||||||| |||||||
Db         11 CCTTCCCTGACCTGGAGTCCGAGTTCCAAGCAGCACTCAGTAGGAAGGTGGCCGAGTTGG 70

Qy        497 TGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTGA 556
              | || ||||| |||||||| |||| | | |  ||||| |||||||||||||| ||||||
Db         71 TTCATTTTCTGCTCCTCAAGTATCGAGCCANGGAGCCGGTCACAAAGGCAGAAATGCTGG 130

Qy        557 CGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAGT 616
              || |||| ||   || |  || | |||||  |||||||| | |||||          |
Db        131 GGAGTGTCGTCGGAAATTGGCAGTATTTCTTTCCTGTGATCTTCAGCAAAGCTTCCAGTT 190

Qy        617 TCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTATT 676
              | |  |||| | |||||||| |  ||||  ||  |||||||  |  |||||  ||
Db        191 CCTTGCAGCTGGTCTTTGGCATCGAGCTGATGGAAGTGGACCCCATCGGCCACTTGTACA 250

Qy        677 TCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATGC 736
              ||||||    |    | | |||| |||| || ||    ||| |||  | |||  |||||
Db        251 TCTTTGCCACCTGCCTGGGCCTCTCCTACGATGGCCTGCTGGGTGACAATCAGATCATGC 310

Qy        737 CCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTCC 796
              |||||    | |||||||  | | ||    || |||   | | ||||  |||| ||
Db        311 CCAAGGCAGGCCTCCTGATAATCGTCCTGGCCATAATCGCAAGAGAGGGCGACTGTGCCC 370

Qy        797 CCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCACT 856
              | ||||||  |||||||| | | ||||||   ||| |||| ||  ||||||||  ||
Db        371 CTGAGGAGAAAATCTGGGAGGAGCTGAGTGTGTTAGAGGTGTTTGAGGGGAGGGAAGACA 430

Qy        857 TTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTGCAGAGAAAGTACCTGG 916
              |||  |  ||||||||||  ||||||||||    |||||  |||||  ||| |||||||
Db        431 GTATCTTGGGGGATCCCAAGAAGCTGCTCACCCAACATTTCGTGCAGGAAAACTACCTGG 490

Qy        917 AGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAATTTTTGTGGGGTCCAAGAG 976
              |||||||| |||| |||  |||| || ||||||||||  ||||||||||||  | |
Db        491 AGTACCGGCAGGTCCCCGGCAGTGATCCTGCATGTTATGAATTCCTGTGGGGTCCAAGGG 550

Qy        977 CCCATTCAGAGGCCAGCAA 995
              |||     ||  |||||  |
Db        551 CCCTCGTTGAAACCAGCTA 569
```

RESULT 13
BQ226610
LOCUS BQ226610 920 bp mRNA linear EST 02-MAY-2002
DEFINITION AGENCOURT_7593822 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6065124 5', mRNA sequence.
ACCESSION BQ226610
VERSION BQ226610.1 GI:20408010
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 920)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13340 row: o column: 13
High quality sequence stop: 587.
FEATURES Location/Qualifiers
source 1. .920
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6065124"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_72"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."

BASE COUNT 202 a 250 c 261 g 207 t
ORIGIN

Query Match 26.0%; Score 271; DB 13; Length 920;
Best Local Similarity 67.8%; Pred. No. 2.3e-54;
Matches 379; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

```
Qy        437 CCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTGG 496
              |||| || ||    || |||  | |     || ||   | ||||||| |||||||
Db         34 CCTTCCCTGACCTGGAGTCCGAGTTCCAAGCAGCACTCAGTAGGAAGGTGGCCGAGTTGG 93

Qy        497 TGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTGA 556
              | || ||||| |||||||| |||| | | |  ||||| |||||||||||||| ||||||
Db         94 TTCATTTTCTGCTCCTCAAGTATCGAGCCAGGGAGCCGGTCACAAAGGCAGAAATGCTGG 153

Qy        557 CGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAGT 616
              || |||| ||   || |  || | |||||  |||||||| | |||||          |
Db        154 GGAGTGTCGTCGGAAATTGGCAGTATTTCTTTCCTGTGATCTTCAGCAAAGCTTCCAGTT 213

Qy        617 TCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTATT 676
              | |  |||| | |||||||| |  ||||  ||  |||||||  |  |||||  ||
Db        214 CCTTGCAGCTGGTCTTTGGCATCGAGCTGATGGAAGTGGACCCCATCGGCCACTTGTACA 273

Qy        677 TCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATGC 736
              ||||||    |    | | |||| |||| || ||    ||| |||  | |||  |||||
Db        274 TCTTTGCCACCTGCCTGGGCCTCTCCTACGATGGCCTGCTGGGTGACAATCAGATCATGC 333

Qy        737 CCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTCC 796
              |||||    | |||||||  | | ||    || |||   | | ||||  |||| ||
Db        334 CCAAGGCAGGCCTCCTGATAATCGTCCTGGCCATAATCGCAAGAGAGGGCGACTGTGCCC 393

Qy        797 CCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCACT 856
              | ||||||  |||||||| | | ||||||   ||| |||| ||  ||||||||  ||
Db        394 CTGAGGAGAAAATCTGGGAGGAGCTGAGTGTGTTAGAGGTGTTTGAGGGGAGGGAAGACA 453

Qy        857 TTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTGCAGAGAAAGTACCTGG 916
              |||  |  ||||||||||  ||||||||||    |||||  |||||  ||| |||||||
Db        454 GTATCTTGGGGGATCCCAAGAAGCTGCTCACCCAACATTTCGTGCAGGAAAACTACCTGG 513

Qy        917 AGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAATTTTTGTGGGGTCCAAGAG 976
              |||||||| |||| |||  |||| || ||||||||||  ||||||||||||  | |
Db        514 AGTACCGGCAGGTCCCCGGCAGTGATCCTGCATGTTATGAATTCCTGTGGGGTCCAAGGG 573

Qy        977 CCCATTCAGAGGCCAGCAA 995
              |||     ||  |||||  |
Db        574 CCCTCGTTGAAACCAGCTA 592
```

```
RESULT 14
BX365439
LOCUS       BX365439               1201 bp    mRNA    linear   EST 05-MAY-2003
DEFINITION  BX365439 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
            Homo sapiens cDNA clone CS0DJ001YA18 5-PRIME, mRNA sequence.
ACCESSION   BX365439
VERSION     BX365439.1  GI:30366911
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 10758.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS1AJ001ZA10QP1&cluster=10758.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS1AJ001ZA10QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ001YA18"
                     /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
                     /cell_line="JURKAT"
                     /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
                     10-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      273 a    304 c    350 g    253 t     21 others
ORIGIN

  Query Match           25.8%;  Score 268.6;  DB 13;  Length 1201;
  Best Local Similarity   66.9%;  Pred. No. 9.3e-54;
  Matches  431;  Conservative    6;  Mismatches  201;  Indels    6;  Gaps    4;

Qy        375 TTTGTGGACCCGATTGGATGAGGAGTCCAGCAGT---GAAGAGGAGGATACAGCTACTTG 431
              ||| |||| || ||| |||||||  |||||||||    ||||||||  ||  ||     | |
Db        500 TTTATGGAGCCAATTCGATGAGGGCTCCAGCAGTYAAGAAGAGGAAGAGCCAAGCTCCTC 559

Qy        432 GCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGA 491
              |   | |    |  |    || | | ||  |    | || ||| |    ||||||||||
Db        560 GGTCGACCCAGCTCAGCTGGAGTTCATGTTCCAAGAAGCACTGAAATTGAAGGTGGCTGA 619

Qy        492 GTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGAT 551
              |||||| || || || |||| |||||||| |   || ||||| ||||||||||||||| ||
Db        620 GTTGGTTCATTTCCTGCTCCACAAATATCGAGTYAAGGAGCCGGTCACAAAGGCAGAAAT 679

Qy        552 GCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCA 611
              ||||  ||  || ||||| || || |||  ||| |||||  | |||||||| ||||||
Db        680 GCTGGAGAGCGTTATCAAAAATTACAAGCGCTACTTTCCTGTTATCTTCGGCAAAGCCTC 739

Qy        612 TGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTC 671
              ||||||||  ||  | || |||||||  ||   |:|  ||  |||||||| |  ||||||
Db        740 CGAGTTCATGCAGGTGATCTTTGGCACTGATGTKAAGGAGGTGGACCCCGCCGGCCACTC 799

Qy        672 CTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGG 731
              |||  || |||       |  | | |||| | |  ||   ||| |  |||||  ||  |
Db        800 CTACATCCTTGTYACTGCTCTTGGCCTCTCGTGCGAWACATGCTGGGGTGATGGTCATAG 859

Qy        732 CATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTG 791
              |||||||||  |  ||||||||  || ||||  || |:||| | |  || | || |||
Db        860 CATGCCCAAGGCCGCCCTCCTGATCATTGTTCTGGGTGTKATCCTAACCAAAGACAACTG 919

Qy        792 TGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAAT-AGGGGTGTGTGCTGGGAGGG 850
              | ||| || ||||| |||||||||| ||||||||  ||  ||||||| || ||||| ||
Db        920 CGCCCCTGAAGAGGTTATCTGGGAAGCGTTGAGTGTKATGGGGGGTGTATGTTGGGAAGG 979

Qy        851 AGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTGCAGAGAAAGT 910
              ||||| |  | || |||||  |||| ||||||||||  | ||||||||||  ||| |
Db        980 AGCACATGTTCTACGGGGA-GCCAGGAAGCTGCTCACCCAAGATTGGGTGCAGGAAAACT 1038

Qy        911 ACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAATTTTTGTGGGGTC 970
              ||||||||| || ||::|||||  ||||  |||| | |  || || |:  |||:||||
Db       1039 ACCTGGAGTYCC-GGMRGTGCCCGGCAGTTATCCTGCGCMCTACGAGTYCCTGTKGGGTT 1097

Qy        971 CAAGAGCCCATTCAGAGGCCAGCAAGAGAAGTCTTAGAGTTTTT 1014
              ||||  ||       | |    | ||||||  |   | || ||
Db       1098 CAAGGCCCACGCTGAAACCAGCTATGAGAAGGGTATAAATTATT 1141


RESULT 15
BG481340
LOCUS       BG481340               820 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602528684F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4652205 5',
            mRNA sequence.
ACCESSION   BG481340
VERSION     BG481340.1  GI:13413619
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 820)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1436  row: g  column: 22
            High quality sequence stop: 798.
FEATURES             Location/Qualifiers
     source          1. .820
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4652205"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_21"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      185 a    215 c    237 g    183 t
ORIGIN

  Query Match           25.5%;  Score 265.8;  DB 10;  Length 820;
  Best Local Similarity   68.3%;  Pred. No. 3.9e-53;
  Matches  427;  Conservative    0;  Mismatches  192;  Indels    6;  Gaps    4;
```

```
Qy        375 TTTGTGGACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAGGATACAGCTACTTG 431
              | |||||| || ||  |||||||  ||||||||   ||||||||||    ||   || |
Db        137 TCTGTGGAGCCAATCCGATGAGGGTTCCAGCAGCAATGAAGAGGAGGGGCCAAGCACCTC 196

Qy        432 GCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGA 491
                |  | |     |  |     || ||| ||  | || | || || ||||| || ||||||||
Db        197 CCCGGACCCAGCTCACCTGGAGTCCCTGTTCCGGGAAGCACTTGATGAGAAAGTGGCTGA 256

Qy        492 GTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGAT 551
              ||| || |  || || |||| ||||||||||  || ||||| ||||||||||||||| ||
Db        257 GTTAGTTCGTTTCCTGCTCCGCAAATATCAAATTAAGGAGCCGGTCACAAAGGCAGAAAT 316

Qy        552 GCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCA 611
              |||   || |||||||||| || || ||| || | |||||    |||||| | ||||||
Db        317 GCTTGAGAGTGTCATCAAAAATTACAAGAACCACTTTCCTGATATCTTCAGCAAAGCCTC 376

Qy        612 TGAGTTCATAGAGCTAATTTTTGGCATTGCCCTG-ACTGATATGGACCCCGACAACCACT 670
              ||||| |||  || | || ||||||||||   || || ||  ||||||| | |  |||||
Db        377 TGAGTGCATGCAGGTGATCTTTGGCATTGATGTGAACGGAAGTGGACCCTGCCGGCCACT 436

Qy        671 CCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGG 730
              |||| || |||    |     | | |||| ||||||| ||     |||  |||||| |||
Db        437 CCTACATCCTTGTCACCTGCCTGGGCCTCTCCTATGATGGCCTGCTGGGTGATGATCAGA 496

Qy        731 GCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCT 790
              | | |||||||| | | |||||||  |  | ||  | |||||||| ||  ||||||||
Db        497 GTACGCCCAAGACCGGCCTCCTGATAATCGTCCTGGGCATGATCTTAATGGAGGGCAGCC 556

Qy        791 GTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCT-GGGAGG 849
              | | ||| |||||||  |||||||||| ||||||||  || ||| ||| || | ||||||
Db        557 GCGCCCCGGAGGAGGCAATCTGGGAAGCGTTGAGTGTGATGGGGCTGTATGATGGGGAGG 616

Qy        850 GAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTGCAGAGAAAG 909
              ||||||  | | ||| || | | ||| ||||||    | | | | |||||||    ||
Db        617 GAGCACAGTGTCTATTGGAAGCTCAGGAAGCTGTCACCCAAAGAGTGGGTGCAGGAGAAC 676

Qy        910 TACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAATTTTTGTGGGGT 969
              |||||||||||||| | ||| || |   |||| ||||   || || |  ||  ||| |||
Db        677 TACCTGGAGTACCCGCAGG-GCGCCGGAGTGATCCTGTGCGCTACGCCTTCCTGTTGGGG 735

Qy        970 CCAAGAGCCCATTCAGAGGCCAGCA 994
              |   ||  |  |  |  |  | ||
Db        736 TCCCAAGGACCCTTTGCTGAAACCA 760
```

Search completed: February 19, 2004, 19:59:13
Job time : 2704 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 16:50:22 ; Search time 404 Seconds
(without alignments)
6955.730 Million cell updates/sec

Title: US-10-085-108-21
Perfect score: 1041
Sequence: 1 ATGCCTCTCTTTCCAAACCT.....CTATCCAGTATCATCCCTAG 1041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1041	100.0	1041	25	ABX95006	cDNA encoding huma
	2	775	74.4	7806	23	AAS88354	DNA encoding novel
	3	486.4	46.7	1983	25	ABX95004	cDNA encoding huma
	4	486.4	46.7	2940	20	AAV69727	Tumour rejection a
	5	486.4	46.7	2940	25	ABX95005	DNA encoding human
	6	484.8	46.6	1983	20	AAV69726	Tumour rejection a
c	7	425	40.8	425	22	ABA46075	Human breast cell
c	8	425	40.8	425	22	ABA56617	Human foetal liver
c	9	425	40.8	425	22	ABA26230	Probe #4696 for ge
c	10	425	40.8	425	22	AAK04758	Human brain expres
c	11	425	40.8	425	22	AAK30282	Human bone marrow
c	12	425	40.8	425	22	AAI14882	Probe #4815 for ge
c	13	425	40.8	425	22	AAI36237	Probe #4923 used t
c	14	425	40.8	425	22	AAI04666	Probe #4657 used t
c	15	425	40.8	425	23	ABS29927	Human liver single
c	16	425	40.8	425	24	ABS04866	Human genome-deriv
c	17	424	40.7	424	22	ABA51188	Human breast cell
c	18	424	40.7	424	22	ABA69192	Human foetal liver
c	19	424	40.7	424	22	ABA36116	Probe #14582 for g
c	20	424	40.7	424	22	AAK17490	Human brain expres
c	21	424	40.7	424	22	AAK43299	Human bone marrow
c	22	424	40.7	424	22	AAI24071	Probe #14004 for g
c	23	424	40.7	424	22	AAI49367	Probe #18053 used
c	24	424	40.7	424	22	AAI09651	Probe #9642 used t
c	25	424	40.7	424	23	ABS42926	Human liver single
c	26	424	40.7	424	24	ABS17378	Human genome-deriv
	27	411.4	39.5	1873	22	ABA57094	Human foetal liver
	28	411.4	39.5	1873	22	ABA26674	Probe #5140 for ge
c	29	411.4	39.5	1963	22	ABA26693	Probe #5159 for ge
	30	411.4	39.5	4031	20	AAV69717	Tumour rejection a
	31	411.4	39.5	4031	25	ABX93690	Human tumour rejec
	32	411.4	39.5	4225	20	AAV69720	Tumour rejection a
	33	411.4	39.5	4225	25	ABX93697	cDNA encoding huma
	34	411.4	39.5	4265	21	AAZ36149	DNA encoding cance
	35	411.4	39.5	4265	23	AAS88353	DNA encoding novel
	36	411.4	39.5	4720	25	ABX95023	DNA encoding human
c	37	411.4	39.5	30274	23	AAS85251	DNA encoding novel
	38	342.2	32.9	1110	25	ABX76240	Lung cancer-associ
	39	342.2	32.9	1528	25	ACC51029	Human bladder canc
	40	342.2	32.9	2559	19	AAV26618	MAGE-10 tumour rej
	41	342.2	32.9	2559	21	AAD01312	Human MAGE-A10 cDN
	42	342.2	32.9	2559	21	AAA52965	Human tumour rejec
	43	342.2	32.9	3510	19	ABQ76210	Human tumour antig
	44	342.2	32.9	3510	21	AAD01311	Human MAGE-A10 gen
	45	340.6	32.7	1545	22	AAF72765	Human prostate can

ALIGNMENTS

RESULT 1
ABX95006
ID ABX95006 standard; cDNA; 1041 BP.
XX
AC ABX95006;
XX
DT 05-JUN-2003 (first entry)
XX
DE cDNA encoding human tumour rejection antigen precursor, MAGE-C3.
XX
KW TRAP; ss; tumour rejection antigen precursor; cytolytic T-cell; CTL;
KW tumour; seminoma; bladder transitional-cell carcinoma; NSCLC; adaptor;
KW head-and-neck squamous-cell carcinoma; breast carcinoma; sarcoma;
KW cutaneous melanoma; nonsmall cell lung cancer; gene; MAGE-C3; human;
KW chromosome Xq27.1-Xq27.3.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1041
FT /*tag= a
FT /product= "MAGE-C3"
XX
PN US2002176865-A1.
XX
PD 28-NOV-2002.
XX
PF 01-MAR-2002; 2002US-0085108.
XX
PR 09-FEB-2000; 2000US-0501104.
PR 25-APR-1997; 97US-0845528.

```
PR   24-APR-1998;   98US-0066281.
PR   17-DEC-1999;   99US-0468433.
XX
PA   (LUCA/) LUCAS S.
PA   (BOON/) BOON-FALLEUR T.
XX
PI   Lucas S,  Boon-Falleur T;
XX
DR   WPI; 2003-328468/31.
DR   P-PSDB; ABU08932.
XX
PT   Novel isolated nucleic acid encoding tumor rejection antigen precursor
PT   MAGE-C3, MAGE-B5, or MAGE-B6, useful as diagnostic probes to determine
PT   presence of abnormal e.g., tumor cells expressing MAGE-C1, MAGE-B5 or
PT   MAGE-B6  -
XX
PS   Claim 1; Fig 5; 59pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a tumour rejection antigen precursor (TRAP) having an amino acid sequence
CC   of a TRAP encoded by a fully defined MAGE-C3, MAGE-B5, or MAGE-B6
CC   polynucleotide sequence. Also disclosed is a method which is useful for
CC   determining presence of cytolytic T-cells specific for complexes of human
CC   leukocyte antigen (HLA) and a peptide derived from the nucleic acid in a
CC   cytotoxic T-lymphocyte (CTL)-containing sample. The nucleic acid is
CC   useful as a diagnostic probe to determine the presence of abnormal
CC   (tumour) cells such as seminoma, bladder transitional-cell carcinoma,
CC   head-and-neck squamous-cell carcinoma, breast carcinoma, sarcoma,
CC   cutaneous melanoma or nonsmall cell lung cancer (NSCLC) which express
CC   MAGE-C1, MAGE-B5 or MAGE-B6. The nucleic acid is useful for diagnosing a
CC   disorder characterised by expression of MAGE-C1, MAGE-B5 or MAGE-B6 TRAPs
CC   or tumour rejection antigens (TRAs). The present sequence represents the
CC   cDNA of the gene encoding the human tumour rejection antigen precursor,
CC   MAGE-C3, which is located on chromosome Xq27.1-Xq27.3.
XX
SQ   Sequence 1041 BP; 242 A; 283 C; 242 G; 274 T; 0 other;

  Query Match           100.0%;  Score 1041;  DB 25;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 2.1e-301;
  Matches 1041;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120

Qy        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180

Qy        181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240

Qy        241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300

Qy        301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360

Qy        361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAGAGGAGGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAGAGGAGGAT 420

Qy        421 ACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAA 480

Qy        481 AAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACA 540

Qy        541 AAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTC 600

Qy        601 GGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCC 660

Qy        661 GACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATT 720

Qy        721 GATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATA 780

Qy        781 AAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGT 840

Qy        841 GCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGGGTG 900

Qy        901 CAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAATTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAATTT 960

Qy        961 TTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAGAAGTCTTAGAGTTTTTATCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAGAAGTCTTAGAGTTTTTATCCAA 1020

Qy       1021 GCTATCCAGTATCATCCCTAG 1041
              |||||||||||||||||||||
Db       1021 GCTATCCAGTATCATCCCTAG 1041
```

```
RESULT 2
AAS88354
ID   AAS88354 standard; cDNA; 7806 BP.
XX
AC   AAS88354;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #24158.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG24167.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
```

```
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 24158; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 7806 BP; 1755 A; 2185 C; 1760 G; 2106 T; 0 other;

  Query Match             74.4%;  Score 775;  DB 23;  Length 7806;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-221;
  Matches  775;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3261 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 3320

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3321 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 3380

Qy        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3381 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 3440

Qy        181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3441 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 3500

Qy        241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3501 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 3560

Qy        301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3561 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 3620

Qy        361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAGAGGAGGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3621 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAGAGGAGGAT 3680

Qy        421 ACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3681 ACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAA 3740

Qy        481 AAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3741 AAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACA 3800

Qy        541 AAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3801 AAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTC 3860

Qy        601 GGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3861 GGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCC 3920

Qy        661 GACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3921 GACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATT 3980

Qy        721 GATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCT 775
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3981 GATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCT 4035


RESULT 3
ABX95004
ID   ABX95004 standard; cDNA; 1983 BP.
XX
AC   ABX95004;
XX
DT   05-JUN-2003  (first entry)
XX
DE   cDNA encoding human tumour rejection antigen precursor, MAGE-C2.
XX
KW   TRAP; ss; tumour rejection antigen precursor; cytolytic T-cell; CTL;
KW   tumour; seminoma; bladder transitional-cell carcinoma; NSCLC; adaptor;
KW   head-and-neck squamous-cell carcinoma; breast carcinoma; sarcoma;
KW   cutaneous melanoma; nonsmall cell lung cancer; gene; MAGE-C2; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             330..1447
FT                   /*tag=  a
FT                   /product= "MAGE-C2"
XX
PN   US2002176865-A1.
XX
PD   28-NOV-2002.
XX
PF   01-MAR-2002; 2002US-0085108.
XX
PR   09-FEB-2000; 2000US-0501104.
PR   25-APR-1997;   97US-0845528.
PR   24-APR-1998;   98US-0066281.
PR   17-DEC-1999;   99US-0468433.
XX
PA   (LUCA/) LUCAS S.
PA   (BOON/) BOON-FALLEUR T.
XX
PI   Lucas S,  Boon-Falleur T;
XX
DR   WPI; 2003-328468/31.
DR   P-PSDB; ABU08931.
XX
PT   Novel isolated nucleic acid encoding tumor rejection antigen precursor
PT   MAGE-C3, MAGE-B5, or MAGE-B6, useful as diagnostic probes to determine
PT   presence of abnormal e.g., tumor cells expressing MAGE-C1, MAGE-B5 or
PT   MAGE-B6  -
XX
PS   Example 11; Page 28-29; 59pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a tumour rejection antigen precursor (TRAP) having an amino acid sequence
CC   of a TRAP encoded by a fully defined MAGE-C3, MAGE-B5, or MAGE-B6
CC   polynucleotide sequence. Also disclosed is a method which is useful for
CC   determining presence of cytolytic T-cells specific for complexes of human
CC   leukocyte antigen (HLA) and a peptide derived from the nucleic acid in a
CC   cytotoxic T-lymphocyte (CTL)-containing sample. The nucleic acid is
CC   useful as a diagnostic probe to determine the presence of abnormal
CC   (tumour) cells such as seminoma, bladder transitional-cell carcinoma,
CC   head-and-neck squamous-cell carcinoma, breast carcinoma, sarcoma,
CC   cutaneous melanoma or nonsmall cell lung cancer (NSCLC) which express
```

```
CC   MAGE-C1, MAGE-B5 or MAGE-B6. The nucleic acid is useful for diagnosing a
CC   disorder characterised by expression of MAGE-C1, MAGE-B5 or MAGE-B6 TRAPs
CC   or tumour rejection antigens (TRAs). The present sequence represents cDNA
CC   encoding the human tumour rejection antigen precursor, MAGE-C2.
XX
SQ   Sequence 1983 BP; 497 A; 428 C; 512 G; 546 T; 0 other;

  Query Match             46.7%;  Score 486.4;  DB 25;  Length 1983;
  Best Local Similarity   73.0%;  Pred. No. 5.9e-135;
  Matches  764;  Conservative    0;  Mismatches  211;  Indels   71;  Gaps    8;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              ||||||| | |||||  | |||||  || || | ||||  | ||||  || || | |   |
Db        330 ATGCCTCCCGTTCCAGGCGTTCCATTCCGCAACGTTGACAACGACTCCCCGACCTCAGTT 389

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              | |  ||| |||| |||||||||||||| || ||| |||||||||  ||||||| ||||||
Db        390 GAGTTAGAAGACTGGGTAGATGCACAGCATCCCACAGATGAGGAAGAGGAGGAAGCCTCC 449

Qy        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
              |||    |||||||| ||||  ||||  ||||  || |||| |||| ||
Db        450 TCC---GCCTCTTCCACTTTGTACTTAGTATTTTCCCCCTCTTCTTTCT---------- 495

Qy        181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
                          ||  | ||||| |||| ||||| |||| ||| || || |||||
Db        496 ----------------CCACATCCTCTTCTCTGATTCTTGGTGGTCCTGAGGAGGAGGAG 539

Qy        241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300
              ||||  ||| || ||| |||  |||| || |||                        |
Db        540 GTGCCCTCTGGTGTGATACCAAATCTTACCGAGA----------------------GC 575

Qy        301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360
              |||||  |   |||| |||||||||||| ||| ||||| |||||||||||||   ||||
Db        576 ATTCCCAGTAGTCCTCCACAGGGTCCTCCACAGGGTCCTTCCCAGAGTCCTCTGAGCTCC 635

Qy        361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAG 417
              ||||  ||| || |||  |||| |  |||  ||||||||||||||      ||| ||||
Db        636 TGCTGCTCCTCTTTTTCATGGAGCTCATTCAGTGAGGAGTCCAGCAGCCAGAAAGGGGAG 695

Qy        418 GATACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGAT 477
              |||||||  || || || | | |||||||  |||||  ||||     ||  ||| | || |||
Db        696 GATACAGGCACCTGTCAGGGCCTGCCAGACAGTGAGTCCTCTTTCACATATACACTAGAT 755

Qy        478 GAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTC 537
              |||||||||||  |||||  ||| |||| || ||||||||||   || || | |||||||
Db        756 GAAAAGGTGGCCGAGTTAGTGGAGTTCCTGCTCCTCAAATACGAAGCAGAGGAGCCTGTA 815

Qy        538 ACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATC 597
              |||  |||||||||||||| || |||||||    |||||  || || || |||||  ||||
Db        816 ACAGAGGCAGAGATGCTGATGATTGTCATC---AAGTACAAAGATTACTTTCCTGTGATA 872

Qy        598 TTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGAC 657
              ||  || ||||| ||||||||| |||||  |||||||| |||||||| ||  ||| |
Db        873 CTCAAGAGAGCCCGTGAGTTCATGGAGCTTCTTTTTGGCCTTGCCCTGATAGAAGTGGGC 932

Qy        658 CCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTG 717
              || |   |||||| || | | |||| | ||||| ||| ||||||| ||||||| ||
Db        933 CCTG---ACCACTTCTGTGTGTTTGCAAACACAGTAGGCCTCACCGATGAGGGTAG---- 985

Qy        718 ATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTC 777
              ||||||  |||||||||||| |||| | |||||||| |||||| ||| ||||||||
Db        986 --TGATGATGAGGGCATGCCCGAGAACAGCCTCCTGATTATTATTCTGAGTGTGATCTTC 1043

Qy        778 ATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTG 837
              ||||||||| ||||| | | |||||||||||||||||||| ||| |||| |||||||
Db       1044 ATAAAGGGCAACTGTGCCTCTGAGGAGGTCATCTGGGAAGTGCTGAATGCAGTAGGGGTA 1103

Qy        838 TGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGG 897
              | |||||||||||||||||  | |||||||| || ||  |||| ||||||| |  ||||
Db       1104 TATGCTGGGAGGGAGCACTTCGTCTATGGGGAGCCTAGGGAGCTCCTCACTAAAGTTTGG 1163

Qy        898 GTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAA 957
              |||||| || | |||||||||||| |||||||||||| ||||| ||||||||  |||||||

Db       1164 GTGCAGGGACATTACCTGGAGTATCGGGAGGTGCCCCACAGTTCTCCTCCATATTATGAA 1223

Qy        958 TTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGA--GAAGTCTTAGAGTTTTTA 1015
              ||  |||||||||||||||||||||||||    || |||||   ||||  |||||||||||
Db       1224 TTCCTGTGGGGTCCAAGAGCCCATTCAGAAAGCATCAAGAAGAAAGTACTAGAGTTTTTA 1283

Qy       1016 TCCAAGCTATCCAGTATCATCCCTAG 1041
              |||||||   ||  |   | |||||
Db       1284 GCCAAGCTGAACAACACTGTTCCTAG 1309
```

RESULT 4
AAV69727
ID AAV69727 standard; DNA; 2940 BP.
XX
AC AAV69727;
XX
DT 01-MAR-1999 (first entry)
XX
DE Tumour rejection antigen precursor MAGE-C2 gene.
XX
KW MAGE-C2; human; tumour rejection antigen precursor; TRAP;
KW therapy; diagnosis; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1287..2408
FT /*tag= a
XX
PN WO9849184-A1.
XX
PD 05-NOV-1998.
XX
PF 24-APR-1998; 98WO-US08493.
XX
PR 25-APR-1997; 97US-0845528.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Boon-Falleur T, De Smet C, Lucas S;
XX
DR WPI; 1999-024041/02.
DR P-PSDB; AAW81547.
XX
PT Tumour rejection antigen precursors - used for determining presence
PT of cytolytic T cells specific for complexes of a human leukocyte
PT antigen
XX
PS Example 11; Page 55-57; 84pp; English.
XX
CC This nucleotide sequence encodes novel human tumour rejection
CC antigen precursor (TRAP) MAGE-C2 (see AAW81547). MAGE-C2 is a novel
CC member of the MAGE family that may be recognised by cytotoxic T
CC cells, leading to lysis of the tumour cells which express it.
CC MACE-C2 and MACE-C1 (see AAW81546) are expressed in a variety of
CC tumours and in normal testis cells, but not by other normal cells.
CC The MAGE-C2 gene was isolated from a melanoma LB373-MEL DNA by PCR
CC using primers based on MAGE-C2 cDNA (see AAV69726). The MACE-C2 gene
CC has been localised to Xq26-q27. It includes at least 2 introns.
CC MACE-C2 and MACE-C1 cDNAs (see AAV69720) are claimed, as are:
CC expression vectors; transformed or transfected cell lines (e.g. COS
CC and CHO); an isolated TRAP encoded by the cDNAs; a kit useful in a
CC PCR based assay; a method for determining expression of a MACE-C1
CC gene using the kit; a polytope comprising a number of tumour
CC rejection antigens derived from MAGE-C1 or MAGE-C2; and a polytope
CC comprising at least one tumour rejection antigen derived from
CC MAGE-C1 or MAGE-C2 and at least one other tumour rejection antigen.
CC MACE-C1 and MACE-C2 can be used in a method for determining the
CC presence of cytolytic T cells specific for complexes of a human
CC leukocyte antigen (HLA).
XX
SQ Sequence 2940 BP; 698 A; 710 C; 801 G; 731 T; 0 other;

```
  Query Match             46.7%;  Score 486.4;  DB 20;  Length 2940;
  Best Local Similarity   73.0%;  Pred. No. 7.4e-135;
  Matches  764;  Conservative    0;  Mismatches  211;  Indels   71;  Gaps    8;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              ||||||| | |||||  | |||||  || || | ||||  | |||| || || | |    |
Db       1287 ATGCCTCCCGTTCCAGGCGTTCCATTCCGCAACGTTGACAACGACTCCCCGACCTCAGTT 1346

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              | |  ||| |||| |||||||||||||| || ||| ||||||||| |||||||| ||||||
Db       1347 GAGTTAGAAGACTGGGTAGATGCACAGCATCCCACAGATGAGGAAGAGGAGGAAGCCTCC 1406

Qy        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
              |||    |||||||| |||| |||| |||| || |||| |||| ||
Db       1407 TCC---GCCTCTTCCACTTTGTACTTAGTATTTTCCCCCTCTTCTTTCT---------- 1452

Qy        181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
                          ||  | ||||| |||| ||||| |||| ||| || || |||||
Db       1453 ----------------CCACATCCTCTTCTCTGATTCTTGGTGGTCCTGAGGAGGAGGAG 1496

Qy        241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300
              ||||  ||| || ||| |||  |||| || |||                         |
Db       1497 GTGCCCTCTGGTGTGATACCAAATCTTACCGAGA----------------------GC 1532

Qy        301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360
              |||||  |   |||| |||||||||||| ||| ||||| ||||||||||||   ||||
Db       1533 ATTCCCAGTAGTCCTCCACAGGGTCCTCCACAGGGTCCTTCCCAGAGTCCTCTGAGCTCC 1592

Qy        361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAG 417
              ||||  ||| || |||  |||| |  |||   |||||||||||||     ||| ||||
Db       1593 TGCTGCTCCTCTTTTTCATGGAGCTCATTCAGTGAGGAGTCCAGCAGCCAGAAAGGGGAG 1652

Qy        418 GATACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGAT 477
              |||||||  || || || | | ||||||| ||||| ||||    ||  ||| | || |||
Db       1653 GATACAGGCACCTGTCAGGGCCTGCCAGACAGTGAGTCCTCTTTCACATATACACTAGAT 1712

Qy        478 GAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTC 537
              |||||||||| ||||| ||| |||| || |||||||||||   || || | ||||||||
Db       1713 GAAAAGGTGGCCGAGTTAGTGGAGTTCCTGCTCCTCAAATACGAAGCAGAGGAGCCTGTA 1772

Qy        538 ACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATC 597
              ||| |||||||||||||| || |||||||   ||||| || || || |||||   ||||
Db       1773 ACAGAGGCAGAGATGCTGATGATTGTCATC---AAGTACAAAGATTACTTTCCTGTGATA 1829

Qy        598 TTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGAC 657
              ||  || ||||| |||||||| |||||  ||||||| |||||||| ||  ||| |
Db       1830 CTCAAGAGAGCCCGTGAGTTCATGGAGCTTCTTTTTGGCCTTGCCCTGATAGAAGTGGGC 1889

Qy        658 CCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTG 717
              || |   |||||| || | | |||| | ||||| ||| ||||||| ||||||| ||
Db       1890 CCTG---ACCACTTCTGTGTGTTTGCAAACACAGTAGGCCTCACCGATGAGGGTAG---- 1942

Qy        718 ATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTC 777
              ||||||  |||||||||| |||| | |||||||| ||||||| ||| ||||||||
Db       1943 --TGATGATGAGGGCATGCCCGAGAACAGCCTCCTGATTATTATTCTGAGTGTGATCTTC 2000

Qy        778 ATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTG 837
              ||||||||| ||||| | | ||||||||||||||||||||| ||| |||| |||||||
Db       2001 ATAAAGGGCAACTGTGCCTCTGAGGAGGTCATCTGGGAAGTGCTGAATGCAGTAGGGGTA 2060

Qy        838 TGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGG 897
              | |||||||||||||||||  | |||||||| || ||  |||| ||||||| |  ||||
Db       2061 TATGCTGGGAGGGAGCACTTCGTCTATGGGGAGCCTAGGGAGCTCCTCACTAAAGTTTGG 2120

Qy        898 GTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAA 957
              |||||| || | |||||||||| |||||||||||| ||||| ||||||| |||||||
Db       2121 GTGCAGGGACATTACCTGGAGTATCGGGAGGTGCCCCACAGTTCTCCTCCATATTATGAA 2180

Qy        958 TTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGA--GAAGTCTTAGAGTTTTTA 1015
              ||  ||||||||||||||||||||||||   || |||||   ||||  ||||||||||
Db       2181 TTCCTGTGGGGTCCAAGAGCCCATTCAGAAAGCATCAAGAAGAAAGTACTAGAGTTTTTA 2240

Qy        1016 TCCAAGCTATCCAGTATCATCCCTAG 1041
               |||||||   ||  |   | |||||
Db        2241 GCCAAGCTGAACAACACTGTTCCTAG 2266
```

```
RESULT 5
ABX95005
ID   ABX95005 standard; DNA; 2940 BP.
XX
AC   ABX95005;
XX
DT   05-JUN-2003  (first entry)
XX
DE   DNA encoding human tumour rejection antigen precursor, MAGE-C2.
XX
KW   TRAP; ds; tumour rejection antigen precursor; cytolytic T-cell; CTL;
KW   tumour; seminoma; bladder transitional-cell carcinoma; NSCLC; adaptor;
KW   head-and-neck squamous-cell carcinoma; breast carcinoma; sarcoma;
KW   cutaneous melanoma; nonsmall cell lung cancer; gene; MAGE-C2; human;
KW   chromosome Xq26-Xq27.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1287..2406
FT                   /*tag=  a
FT                   /product= "MAGE-C2"
XX
PN   US2002176865-A1.
XX
PD   28-NOV-2002.
XX
PF   01-MAR-2002; 2002US-0085108.
XX
PR   09-FEB-2000; 2000US-0501104.
PR   25-APR-1997;   97US-0845528.
PR   24-APR-1998;   98US-0066281.
PR   17-DEC-1999;   99US-0468433.
XX
PA   (LUCA/) LUCAS S.
PA   (BOON/) BOON-FALLEUR T.
XX
PI   Lucas S,  Boon-Falleur T;
XX
DR   WPI; 2003-328468/31.
DR   P-PSDB; ABU08931.
XX
PT   Novel isolated nucleic acid encoding tumor rejection antigen precursor
PT   MAGE-C3, MAGE-B5, or MAGE-B6, useful as diagnostic probes to determine
PT   presence of abnormal e.g., tumor cells expressing MAGE-C1, MAGE-B5 or
PT   MAGE-B6  -
XX
PS   Example 11; Page 30-32; 59pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a tumour rejection antigen precursor (TRAP) having an amino acid sequence
CC   of a TRAP encoded by a fully defined MAGE-C3, MAGE-B5, or MAGE-B6
CC   polynucleotide sequence. Also disclosed is a method which is useful for
CC   determining presence of cytolytic T-cells specific for complexes of human
CC   leukocyte antigen (HLA) and a peptide derived from the nucleic acid in a
CC   cytotoxic T-lymphocyte (CTL)-containing sample. The nucleic acid is
CC   useful as a diagnostic probe to determine the presence of abnormal
CC   (tumour) cells such as seminoma, bladder transitional-cell carcinoma,
CC   head-and-neck squamous-cell carcinoma, breast carcinoma, sarcoma,
CC   cutaneous melanoma or nonsmall cell lung cancer (NSCLC) which express
CC   MAGE-C1, MAGE-B5 or MAGE-B6. The nucleic acid is useful for diagnosing a
CC   disorder characterised by expression of MAGE-C1, MAGE-B5 or MAGE-B6 TRAPs
CC   or tumour rejection antigens (TRAs). The present sequence represents the
CC   human tumour rejection antigen precursor, MAGE-C2, gene which is located
CC   on chromosome Xq26-Xq27.
XX
SQ   Sequence 2940 BP; 698 A; 710 C; 801 G; 731 T; 0 other;
```

```
  Query Match           46.7%;  Score 486.4;  DB 25;  Length 2940;
  Best Local Similarity  73.0%;  Pred. No. 7.4e-135;
  Matches  764;  Conservative    0;  Mismatches  211;  Indels   71;  Gaps    8;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              ||||||| | |||||  | |||||  || || | ||||  | ||||  || || | |    |
Db       1287 ATGCCTCCCGTTCCAGGCGTTCCATTCCGCAACGTTGACAACGACTCCCCGACCTCAGTT 1346

Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
              | |  ||| |||| |||||||||||||| || ||| |||||||| |||||||| ||||||
Db       1347 GAGTTAGAAGACTGGGTAGATGCACAGCATCCCACAGATGAGGAAGAGGAGGAAGCCTCC 1406

Qy        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
              |||    |||||||| ||||  ||||  ||||  || |||| |||| ||
Db       1407 TCC---GCCTCTTCCACTTTGTACTTAGTATTTTCCCCCTCTTCTTTCT----------- 1452

Qy        181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
                          ||  | ||||| |||| |||||  |||| |||  || || |||||
Db       1453 ----------------CCACATCCTCTTCTCTGATTCTTGGTGGTCCTGAGGAGGAGGAG 1496

Qy        241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300
              ||||  ||| || ||| |||  |||| || |||                        |
Db       1497 GTGCCCTCTGGTGTGATACCAAATCTTACCGAGA----------------------GC 1532

Qy        301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360
              |||||  |   |||| ||||||||||| ||| ||||| |||||||||||||    ||||
Db       1533 ATTCCCAGTAGTCCTCCACAGGGTCCTCCACAGGGTCCTTCCCAGAGTCCTCTGAGCTCC 1592

Qy        361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAG 417
              ||||  ||| || |||  |||| |  |||   ||||||||||||||      ||| ||||
Db       1593 TGCTGCTCCTCTTTTTCATGGAGCTCATTCAGTGAGGAGTCCAGCAGCCAGAAAGGGGAG 1652

Qy        418 GATACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGAT 477
              |||||||  || || || | | ||||||| ||||| |||     || ||| | || |||
Db       1653 GATACAGGCACCTGTCAGGGCCTGCCAGACAGTGAGTCCTCTTTCACATATACACTAGAT 1712

Qy        478 GAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTC 537
              |||||||||||| ||||| ||| |||| || ||||||||||  || || | ||||||||
Db       1713 GAAAAGGTGGCCGAGTTAGTGGAGTTCCTGCTCCTCAAATACGAAGCAGAGGAGCCTGTA 1772

Qy        538 ACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATC 597
              ||| ||||||||||||||| || |||||||    ||||| || || || |||||  ||||
Db       1773 ACAGAGGCAGAGATGCTGATGATTGTCATC---AAGTACAAAGATTACTTTCCTGTGATA 1829

Qy        598 TTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGAC 657
              ||  || ||||| ||||||||| |||||  |||||||| |||||||| ||   ||| |
Db       1830 CTCAAGAGAGCCCGTGAGTTCATGGAGCTTCTTTTTGGCCTTGCCCTGATAGAAGTGGGC 1889

Qy        658 CCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTG 717
              || |    |||||| || | | |||| | ||||| ||| ||||||| ||||||| ||
Db       1890 CCTG---ACCACTTCTGTGTGTTTGCAAACACAGTAGGCCTCACCGATGAGGGTAG---- 1942

Qy        718 ATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTC 777
                ||||||  |||||||||||| ||||| | |||||||| ||||||| ||| |||||||
Db       1943 --TGATGATGAGGGCATGCCCGAGAACAGCCTCCTGATTATTATTCTGAGTGTGATCTTC 2000

Qy        778 ATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTG 837
              |||||||||| ||||| | | ||||||||||||||||||||| ||| |||| |||||||
Db       2001 ATAAAGGGCAACTGTGCCTCTGAGGAGGTCATCTGGGAAGTGCTGAATGCAGTAGGGGTA 2060

Qy        838 TGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGG 897
              | |||||||||||||||||  | |||||||| || ||  |||| ||||||| |  ||||
Db       2061 TATGCTGGGAGGGAGCACTTCGTCTATGGGGAGCCTAGGGAGCTCCTCACTAAAGTTTGG 2120

Qy        898 GTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAA 957
              |||||| || | |||||||||||| |||||||||||| ||||| ||||||||  |||||||
Db       2121 GTGCAGGGACATTACCTGGAGTATCGGGAGGTGCCCCACAGTTCTCCTCCATATTATGAA 2180

Qy        958 TTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGA--GAAGTCTTAGAGTTTTTA 1015
              ||  |||||||||||||||||||||||||    || |||||    ||||  ||||||||||
Db       2181 TTCCTGTGGGGTCCAAGAGCCCATTCAGAAAGCATCAAGAAGAAAGTACTAGAGTTTTTA 2240

Qy       1016 TCCAAGCTATCCAGTATCATCCCTAG 1041
```

```
              |||||||   ||  |   | |||||
Db       2241 GCCAAGCTGAACAACACTGTTCCTAG 2266
```

RESULT 6
AAV69726
ID AAV69726 standard; cDNA; 1983 BP.
XX
AC AAV69726;
XX
DT 01-MAR-1999 (first entry)
XX
DE Tumour rejection antigen precursor MAGE-C2 cDNA.
XX
KW MAGE-C2; human; tumour rejection antigen precursor; TRAP;
KW therapy; diagnosis; ds.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   330..1451
FT   /*tag= a
XX
PN WO9849184-A1.
XX
PD 05-NOV-1998.
XX
PF 24-APR-1998; 98WO-US08493.
XX
PR 25-APR-1997; 97US-0845528.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Boon-Falleur T, De Smet C, Lucas S;
XX
DR WPI; 1999-024041/02.
DR P-PSDB; AAW81547.
XX
PT Tumour rejection antigen precursors - used for determining presence
PT of cytolytic T cells specific for complexes of a human leukocyte
PT antigen
XX
PS Claim 1; Page 53-54; 84pp; English.
XX
CC This nucleotide sequence encodes novel human tumour rejection
CC antigen precursor (TRAP) MAGE-C2 (see AAW81547). MAGE-C2 is a novel
CC member of the MAGE family that may be recognised by cytotoxic T
CC cells, leading to lysis of the tumour cells which express it.
CC MACE-C2 and MACE-C1 (see AAW81546) are expressed in a variety of
CC tumours and in normal testis cells, but not by other normal cells.
CC The MAGE-C2 cDNA was isolated from a melanoma LB373-MEL cDNA using
CC a MAGE-C2 probe produced by PCR (see AAV69723-24). The MACE-C2 gene
CC was also isolated (see AAV69727) and has been localised to Xq26-q27.
CC MACE-C2 and MACE-C1 cDNAs (see AAV69720) are claimed, as are:
CC expression vectors; transformed or transfected cell lines (e.g. COS
CC and CHO); an isolated TRAP encoded by the cDNAs; a kit useful in a
CC PCR based assay; a method for determining expression of a MACE-C1
CC gene using the kit; a polytope comprising a number of tumour
CC rejection antigens derived from MAGE-C1 or MAGE-C2; and a polytope
CC comprising at least one tumour rejection antigen derived from
CC MAGE-C1 or MAGE-C2 and at least one other tumour rejection antigen.
CC MACE-C1 and MACE-C2 can be used in a method for determining the
CC presence of cytolytic T cells specific for complexes of a human
CC leukocyte antigen (HLA).
XX
SQ Sequence 1983 BP; 497 A; 428 C; 513 G; 545 T; 0 other;

```
  Query Match           46.6%;  Score 484.8;  DB 20;  Length 1983;
  Best Local Similarity  72.9%;  Pred. No. 1.8e-134;
  Matches  763;  Conservative    0;  Mismatches  212;  Indels   71;  Gaps    8;

Qy          1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60
              ||||||| | |||||  | |||||  || || | ||||  | ||||  || || | |    |
```

```
Db        330 ATGCCTCCCGTTCCAGGCGTTCCATTCCGCAACGTTGACAACGACTCCCCGACCTCAGTT 389
Qy         61 GTGACAGAGGACTTGGTAGATGCACAGGATTCCATAGATGAGGAGGAGGAGGATGCCTCC 120
Db        390 GAGTTAGAAGACTGGGTAGATGCACAGCATCCCACAGATGAGGAAGAGGAGGAAGCCTCC 449
Qy        121 TCCACTTCCTCTTCCTCTTTCCACTTTTTATTCCCCTCCTCCTCTTCCTTGTCCTCATCC 180
Db        450 TCC---GCCTCTTCCACTTTGTACTTAGTATTTTCCCCCTCTTCTTTCT---------- 495
Qy        181 TCACCCTTGTCCTCACCCTTACCCTCTACTCTCATTCTGGGTGTTCCAGAAGATGAGGAT 240
Db        496 ----------------CCACATCCTCTTCTCTGATTCTTGGTGGTCCTGAGGAGGAGGAG 539
Qy        241 ATGCCTGCTGCTGGGATGCCACCTCTTCCCCAGAGTCCTCCTGAGATTCCTCCCCAGGGT 300
Db        540 GTGCCCTCTGGTGTGATACCAAATCTTACCGAGA----------------------GC 575
Qy        301 CCTCCCAAGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCC 360
Db        576 ATTCCCAGTAGTCCTCCACAGGGTCCTCCACAGGGTCCTTCCCAGAGTCCTCTGAGCTCC 635
Qy        361 TGCTCATCCCCTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAG 417
Db        636 TGCTGCTCCTCTTTTTCATGGAGCTCATTCAGTGAGGAGTCCAGCAGCCAGAAAGGGGAG 695
Qy        418 GATACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGAT 477
Db        696 GATACAGGCACCTGTCAGGGCCTGCCAGACAGTGAGTCCTCTTTCACATATACACTAGAT 755
Qy        478 GAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTC 537
Db        756 GAAAAGGTGGCCGAGTTAGTGGAGTTCCTGCTCCTCAAATACGAAGCAGAGGAGCCTGTA 815
Qy        538 ACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATC 597
Db        816 ACAGAGGCAGAGATGCTGATGATTGTCATC---AAGTACAAAGATTACTTTCCTGTGATA 872
Qy        598 TTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGAC 657
Db        873 CTCAAGAGAGCCCGTGAGTTCATGGAGCTTCTTTTTGGCCTTGCCCTGATAGAAGTGGGC 932
Qy        658 CCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTG 717
Db        933 CCTG---ACCACTTCTGTGTGTTTGCAAACACAGTAGGCCTCACCGATGAGGGTAG---- 985
Qy        718 ATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTC 777
Db        986 --TGATGATGAGGGCATGCCCGAGAACAGCCTCCTGATTATTATTCTGAGTGTGATCTTC 1043
Qy        778 ATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTG 837
Db       1044 ATAAAGGGCAACTGTGCCTCTGAGGAGGTCATCTGGGAAGTGCTGAATGCAGTAGGGGTA 1103
Qy        838 TGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGG 897
Db       1104 TATGCTGGGAGGGAGCACTTCGTCTATGGGGAGCCTAGGGAGCTCCTCACTAAAGTTTGG 1163
Qy        898 GTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCCAACAGTGCTCCTCCACGTTATGAA 957
Db       1164 GTGCAGGGACATTACCTGGAGTATCGGGAGGTGCCCCACAGTTCTCCGCCATATTATGAA 1223
Qy        958 TTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGA--GAAGTCTTAGAGTTTTTA 1015
Db       1224 TTCCTGTGGGGTCCAAGAGCCCATTCAGAAAGCATCAAGAAGAAAGTACTAGAGTTTTTA 1283
Qy       1016 TCCAAGCTATCCAGTATCATCCCTAG 1041
Db       1284 GCCAAGCTGAACAACACTGTTCCTAG 1309
```

RESULT 7
ABA46075/c

ID ABA46075 standard; DNA; 425 BP.
XX
AC ABA46075;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #4770.
XX
KW Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
PS Claim 1; SEQ ID NO 4770; 327pp + sequence listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

Query Match 40.8%; Score 425; DB 22; Length 425;
Best Local Similarity 100.0%; Pred. No. 6.6e-117;
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        436 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 495
Db        425 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 366
Qy        496 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 555
Db        365 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 306
Qy        556 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 615
```

```
Db        305 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 246

Qy        616 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 186

Qy        676 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 126

Qy        736 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 795
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 66

Qy        796 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 6

Qy        856 TTTAT 860
              |||||
Db          5 TTTAT 1
```

RESULT 8
ABA56617/c
ID ABA56617 standard; DNA; 425 BP.
XX
AC ABA56617;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #4922.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 4922; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

  Query Match   40.8%; Score 425; DB 22; Length 425;
  Best Local Similarity   100.0%; Pred. No. 6.6e-117;
  Matches 425; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy        436 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 366

Qy        496 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 306

Qy        556 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 246

Qy        616 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 186

Qy        676 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 126

Qy        736 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 795
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 66

Qy        796 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 6

Qy        856 TTTAT 860
              |||||
Db          5 TTTAT 1
```

RESULT 9
ABA26230/c
ID ABA26230 standard; DNA; 425 BP.
XX
AC ABA26230;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #4696 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX

DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 1; SEQ ID No 4696; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

Query Match 40.8%; Score 425; DB 22; Length 425;
Best Local Similarity 100.0%; Pred. No. 6.6e-117;
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      436 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 495
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      425 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 366

Qy      496 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 555
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      365 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 306

Qy      556 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 615
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      305 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 246

Qy      616 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 675
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      245 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 186

Qy      676 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 735
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      185 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 126

Qy      736 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 795
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      125 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 66

Qy      796 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 855
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       65 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 6

Qy      856 TTTAT 860
            |||||
Db        5 TTTAT 1
```

RESULT 10
AAK04758/c
ID AAK04758 standard; DNA; 425 BP.
XX
AC AAK04758;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 4749.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX

OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 4749; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

Query Match 40.8%; Score 425; DB 22; Length 425;
Best Local Similarity 100.0%; Pred. No. 6.6e-117;
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      436 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 495
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      425 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 366

Qy      496 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 555
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      365 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 306

Qy      556 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 615
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      305 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 246

Qy      616 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 675
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      245 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 186

Qy      676 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 735
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      185 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 126

Qy      736 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 795
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      125 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 66

Qy      796 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 855
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       65 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 6

Qy      856 TTTAT 860
            |||||
Db        5 TTTAT 1
```

```
RESULT 11
AAK30282/c
ID   AAK30282 standard; DNA; 425 BP.
XX
AC   AAK30282;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed single exon probe SEQ ID NO: 4839.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
XX
PS   Example 4; SEQ ID NO: 4839; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC   the probes of the invention.
XX
SQ   Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

  Query Match           40.8%;  Score 425;  DB 22;  Length 425;
  Best Local Similarity  100.0%;  Pred. No. 6.6e-117;
  Matches  425;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        436 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 366

Qy        496 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 306

Qy        556 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 246

Qy        616 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 186

Qy        676 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 126

Qy        736 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 795
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 66

Qy        796 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 6

Qy        856 TTTAT 860
              |||||
Db          5 TTTAT 1


RESULT 12
AAI14882/c
ID   AAI14882 standard; DNA; 425 BP.
XX
AC   AAI14882;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #4815 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 4815; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

  Query Match           40.8%;  Score 425;  DB 22;  Length 425;
  Best Local Similarity  100.0%;  Pred. No. 6.6e-117;
  Matches  425;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        436 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        425 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 366

Qy        496 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 306

Qy        556 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 246

Qy        616 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 186

Qy        676 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 126

Qy        736 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 795
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 66

Qy        796 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 6

Qy        856 TTTAT 860
              |||||
Db          5 TTTAT 1
```

RESULT 13
AAI36237/c
ID AAI36237 standard; DNA; 425 BP.
XX
AC AAI36237;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #4923 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 4923; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

```
  Query Match             40.8%;  Score 425;  DB 22;  Length 425;
  Best Local Similarity   100.0%;  Pred. No. 6.6e-117;
  Matches  425;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        436 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 366

Qy        496 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 306

Qy        556 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 246

Qy        616 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 186

Qy        676 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 126

Qy        736 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 795
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 66

Qy        796 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 6

Qy        856 TTTAT 860
              |||||
Db          5 TTTAT 1
```

RESULT 14
AAI04666/c
ID AAI04666 standard; DNA; 425 BP.
XX
AC AAI04666;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #4657 used to measure gene expression in human breast sample.
XX
KW Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.

```
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 25; SEQ ID No 4657; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes.
CC   The present sequence is one such probe. The probes are useful for
CC   measuring human gene expression in a human breast sample, where the probe
CC   hybridises at high stringency to a nucleic acid expressed in the human
CC   breast. The probes are useful for predicting, diagnosing, grading,
CC   staging, monitoring and prognosing diseases of the human breast,
CC   particularly those diseases with polygenic aetiology. The diseases
CC   include: breast cancer, disorders of development, inflammatory diseases
CC   of the breast, fibrocystic changes, proliferative breast disease and
CC   non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

  Query Match             40.8%;  Score 425;  DB 22;  Length 425;
  Best Local Similarity   100.0%;  Pred. No. 6.6e-117;
  Matches  425;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        436 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 366

Qy        496 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 306

Qy        556 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 246

Qy        616 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 186

Qy        676 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 126

Qy        736 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 795
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 66

Qy        796 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 855
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 6

Qy        856 TTTAT 860
              |||||
Db          5 TTTAT 1


RESULT 15
ABS29927/c
ID   ABS29927 standard; DNA; 425 BP.
XX
AC   ABS29927;
XX
DT   25-FEB-2003  (first entry)
XX
DE   Human liver single exon probe, SEQ ID No 4917.
XX
KW   Human; single exon nucleic acid probe; liver; cirrhosis;
KW   hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW   coronary heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157273-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00664.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488898/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analysing gene expression in human adult liver -
XX
PS   Claim 1; SEQ ID No 4917; 658pp; English.
XX
CC   The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC   measuring human gene expression in a sample derived from human adult
CC   liver, comprising one of 13109 defined nucleotide sequences given in the
CC   specification (or complements/ fragments). The probe hybridises at high
CC   stringency to a nucleic acid molecule expressed in the human adult
CC   liver. (I) may be used for predicting, measuring and displaying gene
CC   expression in samples derived from human adult liver. The genes
CC   identified may be involved in genetic liver diseases such as cirrhosis,
CC   hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
CC   is associated with coronary heart disease. ABS25011-ABS51005 represent
CC   human liver single exon nucleic acid probes of the invention.
CC   Note: The sequence information for this patent does not appear in the
CC   printed specification but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

  Query Match             40.8%;  Score 425;  DB 23;  Length 425;
  Best Local Similarity   100.0%;  Pred. No. 6.6e-117;
  Matches  425;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        436 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 GCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTG 366

Qy        496 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 GTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTCACAAAGGCAGAGATGCTG 306

Qy        556 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 ACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATCTTCGGGAAAGCCCATGAG 246

Qy        616 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 TTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGACCCCGACAACCACTCCTAT 186

Qy        676 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 TTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTGATTGATGACCAGGGCATG 126

Qy        736 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 795
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          125 CCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTCATAAAGGGCAGCTGTGTC 66

Qy          796 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 855
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           65 CCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCAC 6

Qy          856 TTTAT 860
                |||||
Db            5 TTTAT 1
```

Search completed: February 19, 2004, 18:04:09
Job time : 408 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 18:04:16 ; Search time 97 Seconds
(without alignments)
4736.906 Million cell updates/sec

Title: US-10-085-108-21
Perfect score: 1041
Sequence: 1 ATGCCTCTCTTTCCAAACCT.....CTATCCAGTATCATCCCTAG 1041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

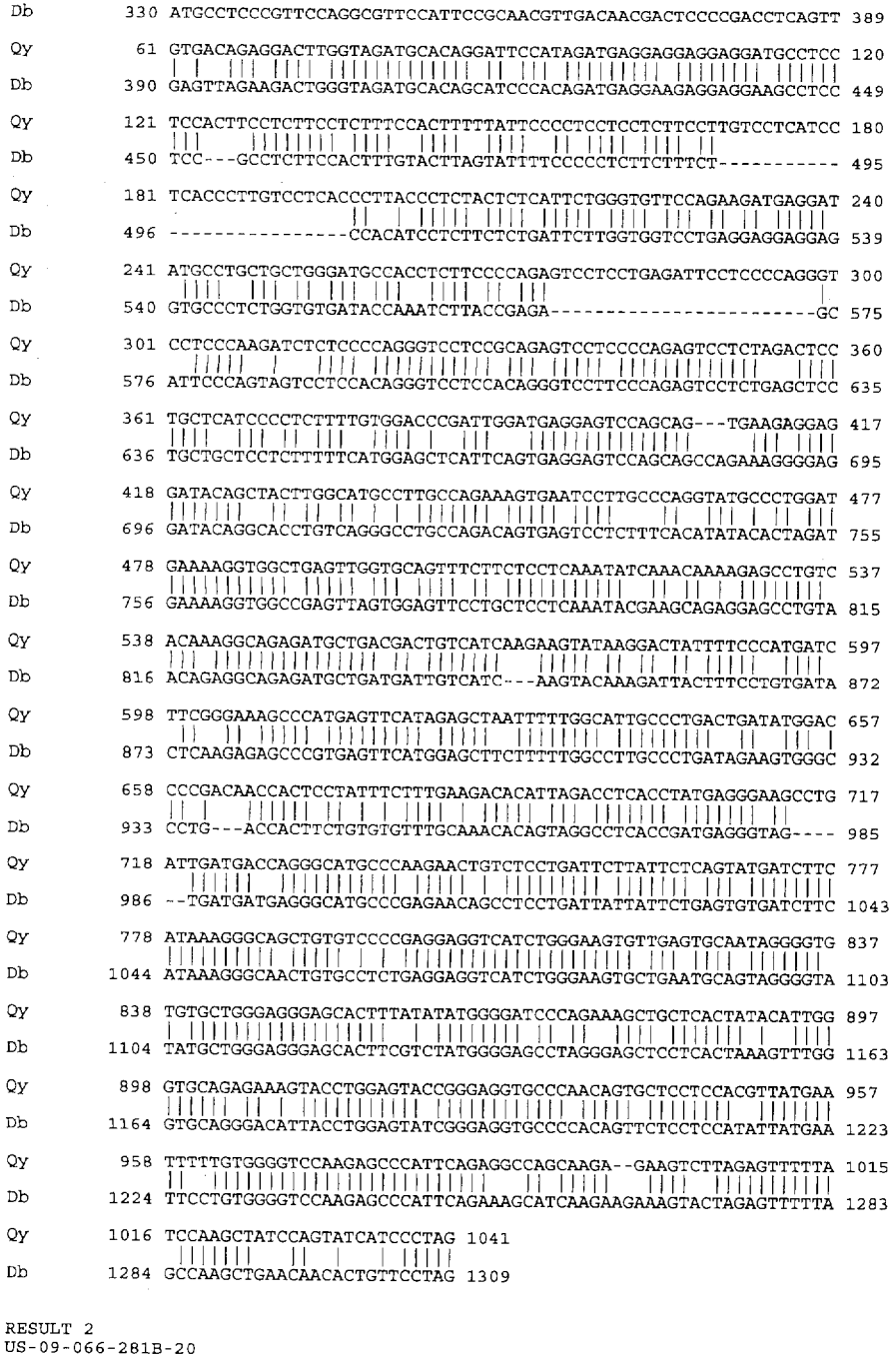
Result No.	Score	% Query Match	Length	DB	ID	Description
1	486.4	46.7	1983	4	US-09-066-281B-18	Sequence 18, Appl
2	486.4	46.7	2940	4	US-09-066-281B-20	Sequence 20, Appl
3	411.4	39.5	4031	2	US-08-993-118-1	Sequence 1, Appli
4	411.4	39.5	4031	3	US-08-845-528C-1	Sequence 1, Appli
5	411.4	39.5	4031	4	US-09-066-281B-1	Sequence 1, Appli
6	411.4	39.5	4225	2	US-08-993-118-9	Sequence 9, Appli
7	411.4	39.5	4225	3	US-08-845-528C-9	Sequence 9, Appli
8	411.4	39.5	4225	4	US-09-066-281B-9	Sequence 9, Appli
9	411.4	39.5	4265	3	US-09-061-709-1	Sequence 1, Appli
10	411.4	39.5	4265	4	US-09-899-651-1	Sequence 1, Appli
11	342.2	32.9	2559	2	US-08-724-774B-3	Sequence 3, Appli
12	342.2	32.9	2559	3	US-09-089-595-3	Sequence 3, Appli
13	342.2	32.9	2559	3	US-09-382-855-3	Sequence 3, Appli
14	342.2	32.9	2559	3	US-09-183-714B-3	Sequence 3, Appli
15	342.2	32.9	2559	4	US-09-642-281-3	Sequence 3, Appli
16	342.2	32.9	2559	4	US-09-589-717-3	Sequence 3, Appli
17	342.2	32.9	3510	3	US-09-056-105-16	Sequence 16, Appl
18	314.4	30.2	1412	1	US-08-299-849B-21	Sequence 21, Appl
19	314.4	30.2	1412	2	US-08-142-368A-21	Sequence 21, Appl
20	314.4	30.2	1412	3	US-08-967-727-21	Sequence 21, Appl
21	314.4	30.2	1412	3	US-08-037-230D-21	Sequence 21, Appl
22	314.4	30.2	1412	4	US-09-583-850-21	Sequence 21, Appl
23	314.4	30.2	1412	4	US-09-579-197-21	Sequence 21, Appl
24	314.4	30.2	1412	4	US-09-404-026-21	Sequence 21, Appl
25	314.4	30.2	2931	3	US-09-056-105-15	Sequence 15, Appl
26	312.2	30.0	1810	1	US-08-299-849B-20	Sequence 20, Appl
27	312.2	30.0	1810	2	US-08-142-368A-20	Sequence 20, Appl
28	312.2	30.0	1810	3	US-08-967-727-20	Sequence 20, Appl
29	312.2	30.0	1810	3	US-08-037-230D-20	Sequence 20, Appl
30	312.2	30.0	1810	4	US-09-583-850-20	Sequence 20, Appl
31	312.2	30.0	1810	4	US-09-579-197-20	Sequence 20, Appl
32	312.2	30.0	1810	4	US-09-404-026-20	Sequence 20, Appl
33	312.2	30.0	3839	3	US-09-056-105-14	Sequence 14, Appl
34	310.2	29.8	3672	3	US-09-056-105-17	Sequence 17, Appl
35	303	29.1	1107	1	US-08-299-849B-23	Sequence 23, Appl
36	303	29.1	1107	2	US-08-142-368A-23	Sequence 23, Appl
37	303	29.1	1107	3	US-08-967-727-23	Sequence 23, Appl
38	303	29.1	1107	3	US-08-037-230D-23	Sequence 23, Appl
39	303	29.1	1107	4	US-09-583-850-23	Sequence 23, Appl
40	303	29.1	1107	4	US-09-404-026-23	Sequence 23, Appl
41	298.6	28.7	1691	2	US-08-993-118-8	Sequence 8, Appli
42	298.6	28.7	1691	3	US-08-845-528C-8	Sequence 8, Appli
43	298.6	28.7	1691	4	US-09-066-281B-8	Sequence 8, Appli
44	298.6	28.7	2420	1	US-08-465-167A-23	Sequence 23, Appl
45	298.6	28.7	2420	3	US-09-056-105-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-066-281B-18
; Sequence 18, Application US/09066281B
; Patent No. 6475783
; GENERAL INFORMATION:
; APPLICANT: LUCAS, Sophie; DE SMET, Charles; BOON-FALLEUR, Thierry
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING
; TITLE OF INVENTION: FOR TUMOR REJECTION ANTIGEN PRECURSOR MAGE-C1 AND MAGE-C2
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,281B
; FILING DATE: April 24, 1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/845,528
; FILING DATE: April 25, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5455.2 US - JEL/MAS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 318-3100
; TELEFAX: (212) 752-5958
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1983 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double-stranded
; TOPOLOGY: linear
US-09-066-281B-18

Query Match 46.7%; Score 486.4; DB 4; Length 1983;
Best Local Similarity 73.0%; Pred. No. 1e-116;
Matches 764; Conservative 0; Mismatches 211; Indels 71; Gaps 8;

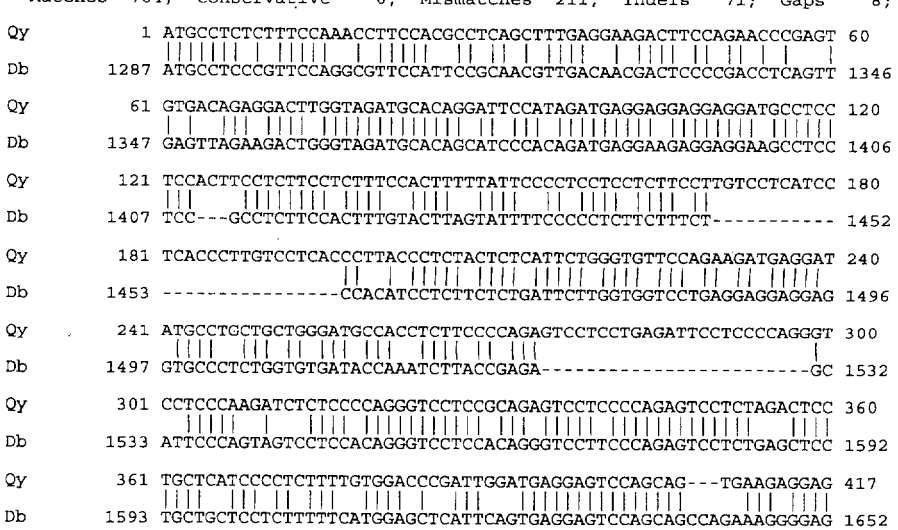
Qy 1 ATGCCTCTCTTTCCAAACCTTCCACGCCTCAGCTTTGAGGAAGACTTCCAGAACCCGAGT 60



RESULT 2
US-09-066-281B-20

; Sequence 20, Application US/09066281B
; Patent No. 6475783
; GENERAL INFORMATION:
; APPLICANT: LUCAS, Sophie; DE SMET, Charles; BOON-FALLEUR, Thierry
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING
; TITLE OF INVENTION: FOR TUMOR REJECTION ANTIGEN PRECURSOR MAGE-C1 AND MAGE-C2
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,281B
; FILING DATE: April 24, 1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/845,528
; FILING DATE: April 25, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5455.2 US - JEL/MAS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 318-3100
; TELEFAX: (212) 752-5958
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2940 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double-stranded
; TOPOLOGY: linear
US-09-066-281B-20

Query Match 46.7%; Score 486.4; DB 4; Length 2940;
Best Local Similarity 73.0%; Pred. No. 1.2e-116;
Matches 764; Conservative 0; Mismatches 211; Indels 71; Gaps 8;



```
Qy        418 GATACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATGCCCTGGAT 477
Db       1653 GATACAGGCACCTGTCAGGGCCTGCCAGACAGTGAGTCCTCTTTCACATATACACTAGAT 1712

Qy        478 GAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAGAGCCTGTC 537
Db       1713 GAAAAGGTGGCCGAGTTAGTGGAGTTCCTGCTCCTCAAATACGAAGCAGAGGAGCCTGTA 1772

Qy        538 ACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTCCCATGATC 597
Db       1773 ACAGAGGCAGAGATGCTGATGATTGTCATC---AAGTACAAAGATTACTTTCCTGTGATA 1829

Qy        598 TTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTGATATGGAC 657
Db       1830 CTCAAGAGAGCCCGTGAGTTCATGGAGCTTCTTTTTGGCCTTGCCCTGATAGAAGTGGGC 1889

Qy        658 CCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGGGAAGCCTG 717
Db       1890 CCTG---ACCACTTCTGTGTGTTTGCAAACACAGTAGGCCTCACCGATGAGGGTAG---- 1942

Qy        718 ATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTATGATCTTC 777
Db       1943 --TGATGATGAGGGCATGCCCGAGAACAGCCTCCTGATTATTATTCTGAGTGTGATCTTC 2000

Qy        778 ATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAATAGGGGTG 837
Db       2001 ATAAAGGGCAACTGTGCCTCTGAGGAGGTCATCTGGGAAGTGCTGAATGCAGTAGGGGTA 2060

Qy        838 TGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTATACATTGG 897
Db       2061 TATGCTGGGAGGGAGCACTTCGTCTATGGGGAGCCTAGGGAGCTCCTCACTAAAGTTTGG 2120

Qy        898 GTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCACGTTATGAA 957
Db       2121 GTGCAGGGACATTACCTGGAGTATCGGGAGGTGCCCCACAGTTCTCCTCCATATTATGAA 2180

Qy        958 TTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGA--GAAGTCTTAGAGTTTTTA 1015
Db       2181 TTCCTGTGGGGTCCAAGAGCCCATTCAGAAAGCATCAAGAAGAAAGTACTAGAGTTTTTA 2240

Qy       1016 TCCAAGCTATCCAGTATCATCCCTAG 1041
Db       2241 GCCAAGCTGAACAACACTGTTCCTAG 2266
```

```
RESULT 3
US-08-993-118-1
; Sequence 1, Application US/08993118
; Patent No. 5997872
;  GENERAL INFORMATION:
;    APPLICANT:  LUCAS, Sophie;
;    APPLICANT:  DE SMET, Charles;
;    APPLICANT:  BOON-FALLEUR, Thierry
;    TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
;    TITLE OF INVENTION:  REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 360 kb storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/993,118
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/845,528
;      FILING DATE:  April 25, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mary Anne Schofield
;      REGISTRATION NUMBER:  36,669
;      REFERENCE/DOCKET NUMBER:  LUD 5455
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4031 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double-stranded
;      TOPOLOGY:  linear
US-08-993-118-1

  Query Match             39.5%;  Score 411.4;  DB 2;  Length 4031;
  Best Local Similarity   74.1%;  Pred. No. 3.6e-97;
  Matches  608;  Conservative    0;  Mismatches  166;  Indels   47;  Gaps    5;

Qy        264 TCTTCCCCAGAGTCCTCCTGAGATTCC---TCCCCAGGGTCCTCCCAAGATCTCTCCCCA 320
Db       2518 TCTTTCCCAGAGTTCTCCTGTGAGCTCCTTCCCCTCCTCCACTTCATCGAGTCTTTCCAA 2577

Qy        321 GGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCATCCCCTCTTTTGTG 380
Db       2578 GAGTTCCCCTGAGAGTCCTCTCCAGAGTCCTGTGATCTCCTTCTCCTCCTCCACTTCATT 2637

Qy        381 GACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAGGATACAGCTACTTGGCATGC 437
Db       2638 GAGCCCATTCAGTGAAGAGTCCAGCAGCCCAGTAGATGAATATACAAGTTCCTCAGACAC 2697

Qy        438 CTTGCCAGAAAGTGAATCCTTGCC---------------------------------- 461
Db       2698 CTTGCTAGAGAGTGATTCCTTGACAGACAGCGAGTCCTTGATAGAGAGCGAGCCCTTGTT 2757

Qy        462 CAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCA 521
Db       2758 CACTTATACACTGGATGAAAAGGTGGACGAGTTGGCGCGGTTTCTTCTCCTCAAATATCA 2817

Qy        522 AACAAAAGAGCCTGTCACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGA 581
Db       2818 AGTGAAGCAGCCTATCACAAAGGCAGAGATGCTGACGAATGTCATCAGCAGGTACACGGG 2877

Qy        582 CTATTTTCCCATGATCTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGC 641
Db       2878 CTACTTTCCTGTGATCTTCAGGAAAGCCCGTGAGTTCATAGAGATACTTTTTGGCATTTC 2937

Qy        642 CCTGACTGATATGGACCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCAC 701
Db       2938 CCTGAGAGAAGTGGACCCTG---ATGACTCCTATGTCTTTGTAAACACATTAGACCTCAC 2994

Qy        702 CTATGAGGGAAGCCTGATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTAT 761
Db       2995 CTCTGAGGGGTGTCTGAGTGATGAGCAGGGCATGTCCCAGAACCGCCTCCTGATTCTTAT 3054

Qy        762 TCTCAGTATGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTT 821
Db       3055 TCTGAGTATCATCTTCATAAAGGGCACCTATGCCTCTGAGGAGGTCATCTGGGATGTGCT 3114

Qy        822 GAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCT 881
Db       3115 GAGTGGAATAGGGGTGCGTGCTGGGAGGGAGCACTTTGCCTTTGGGGAGCCCAGGGAGCT 3174

Qy        882 GCTCACTATACATTGGGTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGC 941
Db       3175 CCTCACTAAAGTTTGGGTGCAGGAACATTACCTAGAGTACCGGGAGGTGCCCAACTCTTC 3234

Qy        942 TCCTCCACGTTATGAATTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAG--A 999
```

```
Db       3235 TCCTCCTCGTTACGAATTCCTGTGGGGTCCAAGAGCTCATTCAGAAGTCATTAAGAGGAA 3294

Qy       1000 AGTCTTAGAGTTTTTATCCAAGCTATCCAGTATCATCCCTA 1040
              ||| |||||||||| ||| |||| | || | ||||||
Db       3295 AGTAGTAGAGTTTTTGGCCATGCTAAAGAATACCGTCCCTA 3335


RESULT 4
US-08-845-528C-1
; Sequence 1, Application US/08845528C
; Patent No. 6027924
;  GENERAL INFORMATION:
;    APPLICANT:  LUCAS, Sophie;
;    APPLICANT:  DE SMET, Charles;
;    APPLICANT:  BOON-FALLEUR, Thierry
;    TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
;    TITLE OF INVENTION:  REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 360 kb storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/845,528C
;      FILING DATE:  April 25, 1997
;      CLASSIFICATION:  4335
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mary Anne Schofield
;      REGISTRATION NUMBER:  36,669
;      REFERENCE/DOCKET NUMBER:  LUD 5455
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4031 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double-stranded
;      TOPOLOGY:  linear
US-08-845-528C-1

  Query Match             39.5%;  Score 411.4;  DB 3;  Length 4031;
  Best Local Similarity   74.1%;  Pred. No. 3.6e-97;
  Matches  608;  Conservative    0;  Mismatches  166;  Indels   47;  Gaps    5;

Qy        264 TCTTCCCCAGAGTCCTCCTGAGATTCC---TCCCCAGGGTCCTCCCAAGATCTCTCCCCA 320
              |||| |||||||| |||||| ||   |     |||       || |   ||    | ||| |
Db       2518 TCTTTCCCAGAGTTCTCCTGTGAGCTCCTTCCCCTCCTCCACTTCATCGAGTCTTTCCAA 2577

Qy        321 GGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCATCCCCTCTTTTGTG 380
              | || | ||  |||||||| |||||||||| |    ||||| ||| ||| |   ||  |
Db       2578 GAGTTCCCCTGAGAGTCCTCTCCAGAGTCCTGTGATCTCCTTCTCCTCCTCCACTTCATT 2637

Qy        381 GACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAGGATACAGCTACTTGGCATGC 437
              || || |||   ||| |||||||||||     | ||| ||  |||||  | | |   | |
Db       2638 GAGCCCATTCAGTGAAGAGTCCAGCAGCCCAGTAGATGAATATACAAGTTCCTCAGACAC 2697

Qy        438 CTTGCCAGAAAGTGAATCCTTGCC---------------------------------- 461
              ||||| ||| ||||| |||||| |
Db       2698 CTTGCTAGAGAGTGATTCCTTGACAGACAGCGAGTCCTTGATAGAGAGCGAGCCCTTGTT 2757

Qy        462 CAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCA 521
              ||  ||| | |||||||||||||||  ||||||| || ||||||||||||||||||||
Db       2758 CACTTATACACTGGATGAAAAGGTGGACGAGTTGGCGCGGTTTCTTCTCCTCAAATATCA 2817

Qy        522 AACAAAAGAGCCTGTCACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGA 581
              |   ||  |||||  |||||||||||||||||||||||||| ||||||||  | ||| | ||
Db       2818 AGTGAAGCAGCCTATCACAAAGGCAGAGATGCTGACGAATGTCATCAGCAGGTACACGGG 2877

Qy        582 CTATTTTCCCATGATCTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGC 641
              ||| |||||  |||||||| |||||||| |||||||||||||  || ||||||||||| |
Db       2878 CTACTTTCCTGTGATCTTCAGGAAAGCCCGTGAGTTCATAGAGATACTTTTTGGCATTTC 2937

Qy        642 CCTGACTGATATGGACCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCAC 701
              |||||  ||  |||||||  |   | |||||||| |||||| | ||||||||||||||||
Db       2938 CCTGAGAGAAGTGGACCCTG---ATGACTCCTATGTCTTTGTAAACACATTAGACCTCAC 2994

Qy        702 CTATGAGGGAAGCCTGATTGATGACCAGGGCATGCCCCAAGAACTGTCTCCTGATTCTTAT 761
              ||  ||||||  | |||| |||||| |||||||| || ||||| | ||||||||||||||
Db       2995 CTCTGAGGGGTGTCTGAGTGATGAGCAGGGCATGTCCCAGAACCGCCTCCTGATTCTTAT 3054

Qy        762 TCTCAGTATGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTT 821
              ||| ||||| |||||||||||||||| || || | | |||||||||||||||||| ||| |
Db       3055 TCTGAGTATCATCTTCATAAAGGGCACCTATGCCTCTGAGGAGGTCATCTGGGATGTGCT 3114

Qy        822 GAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCT 881
              ||||| |||||||||| |||||||||||||||||||    | |||||| |||| ||||
Db       3115 GAGTGGAATAGGGGTGCGTGCTGGGAGGGAGCACTTTGCCTTTGGGGAGCCCAGGGAGCT 3174

Qy        882 GCTCACTATACATTGGGTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGC 941
              |||||||  |  |||||||||  |  | |||| |||||||||||||||||||  | |
Db       3175 CCTCACTAAAGTTTGGGTGCAGGAACATTACCTAGAGTACCGGGAGGTGCCCAACTCTTC 3234

Qy        942 TCCTCCACGTTATGAATTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAG--A 999
              |||||| |||||  |||||  |||||||||||||||| |||||||  | ||  |||||  |
Db       3235 TCCTCCTCGTTACGAATTCCTGTGGGGTCCAAGAGCTCATTCAGAAGTCATTAAGAGGAA 3294

Qy       1000 AGTCTTAGAGTTTTTATCCAAGCTATCCAGTATCATCCCTA 1040
              |||  ||||||||||  |||  ||||    | || | ||||||
Db       3295 AGTAGTAGAGTTTTTGGCCATGCTAAAGAATACCGTCCCTA 3335


RESULT 5
US-09-066-281B-1
; Sequence 1, Application US/09066281B
; Patent No. 6475783
;  GENERAL INFORMATION:
;    APPLICANT:  LUCAS, Sophie; DE SMET, Charles; BOON-FALLEUR, Thierry
;    TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE CODING
;    TITLE OF INVENTION:  FOR TUMOR REJECTION ANTIGEN PRECURSOR MAGE-C1 AND MAGE-C2
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fulbright & Jaworski L.L.P.
;      STREET:  666 Fifth Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 360 kb storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/066,281B
;      FILING DATE:  April 24, 1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/845,528
;      FILING DATE:  April 25, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mary Anne Schofield
;      REGISTRATION NUMBER:  36,669
;      REFERENCE/DOCKET NUMBER:  LUD 5455.2 US - JEL/MAS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 318-3100
```

```
;       TELEFAX:  (212) 752-5958
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  4031 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double-stranded
;       TOPOLOGY:  linear
US-09-066-281B-1

  Query Match             39.5%;  Score 411.4;  DB 4;  Length 4031;
  Best Local Similarity   74.1%;  Pred. No. 3.6e-97;
  Matches  608;  Conservative    0;  Mismatches  166;  Indels   47;  Gaps    5;

Qy        264 TCTTCCCCAGAGTCCTCCTGAGATTCC---TCCCCAGGGTCCTCCCAAGATCTCTCCCCA 320
              |||| ||||||||  ||||||  ||  |  |||     || |   ||    | || |
Db       2518 TCTTTCCCAGAGTTCTCCTGTGAGCTCCTTCCCCTCCTCCACTTCATCGAGTCTTTCCAA 2577

Qy        321 GGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCATCCCCTCTTTTGTG 380
              | ||  | ||  |||||||||| ||||||||||  |    |||||  |||  |||  |    ||  |
Db       2578 GAGTTCCCCTGAGAGTCCTCTCCAGAGTCCTGTGATCTCCTTCTCCTCCTCCACTTCATT 2637

Qy        381 GACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAGGATACAGCTACTTGGCATGC 437
              ||  ||  |||    |||  ||||||||||||     | |||  ||  |||||  | | |    | |
Db       2638 GAGCCCATTCAGTGAAGAGTCCAGCAGCCCAGTAGATGAATATACAAGTTCCTCAGACAC 2697

Qy        438 CTTGCCAGAAAGTGAATCCTTGCC---------------------------------- 461
              |||||  |||  |||||  ||||||  |
Db       2698 CTTGCTAGAGAGTGATTCCTTGACAGACAGCGAGTCCTTGATAGAGAGCGAGCCCTTGTT 2757

Qy        462 CAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCA 521
              ||   |||  |  ||||||||||||||||  |||||||  ||  ||||||||||||||||||
Db       2758 CACTTATACACTGGATGAAAAGGTGGACGAGTTGGCGCGGTTTCTTCTCCTCAAATATCA 2817

Qy        522 AACAAAAGAGCCTGTCACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGA 581
              |    ||  |||||  |||||||||||||||||||||||||  ||||||||  | |||  | ||
Db       2818 AGTGAAGCAGCCTATCACAAAGGCAGAGATGCTGACGAATGTCATCAGCAGGTACACGGG 2877

Qy        582 CTATTTTCCCATGATCTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGC 641
              |||  |||||   |||||||  ||||||||  ||||||||||||  ||  ||||||||||| |
Db       2878 CTACTTTCCTGTGATCTTCAGGAAAGCCCGTGAGTTCATAGAGATACTTTTTGGCATTTC 2937

Qy        642 CCTGACTGATATGGACCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCAC 701
              |||||  ||  ||||||||  |  ||||||||  ||||||  |  ||||||||||||||
Db       2938 CCTGAGAGAAGTGGACCCTG---ATGACTCCTATGTCTTTGTAAACACATTAGACCTCAC 2994

Qy        702 CTATGAGGGAAGCCTGATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTAT 761
              ||  |||||||  |  ||||  ||||||  |||||||||  ||  |||||  |  |||||||||||||
Db       2995 CTCTGAGGGGTGTCTGAGTGATGAGCAGGGCATGTCCCAGAACCGCCTCCTGATTCTTAT 3054

Qy        762 TCTCAGTATGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTT 821
              |||  |||||  ||||||||||||||||||  ||  ||  |  |  ||||||||||||||||||  |||  |
Db       3055 TCTGAGTATCATCTTCATAAAGGGCACCTATGCCTCTGAGGAGGTCATCTGGGATGTGCT 3114

Qy        822 GAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCT 881
              |||||  ||||||||||  ||||||||||||||||||||  |  ||||||  |||||  ||||
Db       3115 GAGTGGAATAGGGGTGCGTGCTGGGAGGGAGCACTTTGCCTTTGGGGAGCCCAGGGAGCT 3174

Qy        882 GCTCACTATACATTGGGTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGC 941
              |||||||  |  ||||||||||  |  |  |||||  |||||||||||||||||||||  | |
Db       3175 CCTCACTAAAGTTTGGGTGCAGGAACATTACCTAGAGTACCGGGAGGTGCCCAACTCTTC 3234

Qy        942 TCCTCCACGTTATGAATTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAG--A 999
              ||||||  |||||  ||||  ||||||||||||||||  |||||||  |  ||  |||||  |
Db       3235 TCCTCCTCGTTACGAATTCCTGTGGGGTCCAAGAGCTCATTCAGAAGTCATTAAGAGGAA 3294

Qy       1000 AGTCTTAGAGTTTTTATCCAAGCTATCCAGTATCATCCCTA 1040
              |||  ||||||||||  |||  ||||  |  ||  |  ||||||
Db       3295 AGTAGTAGAGTTTTTGGCCATGCTAAAGAATACCGTCCCTA 3335


RESULT 6
US-08-993-118-9
; Sequence 9, Application US/08993118
; Patent No. 5997872
;  GENERAL INFORMATION:
;    APPLICANT:  LUCAS, Sophie;
;    APPLICANT:  DE SMET, Charles;
;    APPLICANT:  BOON-FALLEUR, Thierry
;    TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
;    TITLE OF INVENTION:  REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 360 kb storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/993,118
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/845,528
;      FILING DATE:  April 25, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mary Anne Schofield
;      REGISTRATION NUMBER:  36,669
;      REFERENCE/DOCKET NUMBER:  LUD 5455
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4225 base pairs
;      TYPE:  nucleic acids
;      STRANDEDNESS:  double-stranded
;      TOPOLOGY:  linear
US-08-993-118-9

  Query Match             39.5%;  Score 411.4;  DB 2;  Length 4225;
  Best Local Similarity   74.1%;  Pred. No. 3.6e-97;
  Matches  608;  Conservative    0;  Mismatches  166;  Indels   47;  Gaps    5;

Qy        264 TCTTCCCCAGAGTCCTCCTGAGATTCC---TCCCCAGGGTCCTCCCAAGATCTCTCCCCA 320
              |||| ||||||||  ||||||  ||  |  |||     || |   ||    | || |
Db       2728 TCTTTCCCAGAGTTCTCCTGTGAGCTCCTTCCCCTCCTCCACTTCATCGAGTCTTTCCAA 2787

Qy        321 GGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCATCCCCTCTTTTGTG 380
              | ||  | ||  |||||||||| ||||||||||  |    |||||  |||  |||  |    ||  |
Db       2788 GAGTTCCCCTGAGAGTCCTCTCCAGAGTCCTGTGATCTCCTTCTCCTCCTCCACTTCATT 2847

Qy        381 GACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAGGATACAGCTACTTGGCATGC 437
              ||  ||  |||    |||  ||||||||||||     | |||  ||  |||||  | | |    | |
Db       2848 GAGCCCATTCAGTGAAGAGTCCAGCAGCCCAGTAGATGAATATACAAGTTCCTCAGACAC 2907

Qy        438 CTTGCCAGAAAGTGAATCCTTGCC---------------------------------- 461
              |||||  |||  |||||  ||||||  |
Db       2908 CTTGCTAGAGAGTGATTCCTTGACAGACAGCGAGTCCTTGATAGAGAGCGAGCCCTTGTT 2967

Qy        462 CAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCA 521
              ||   |||  |  ||||||||||||||||  |||||||  ||  ||||||||||||||||||
Db       2968 CACTTATACACTGGATGAAAAGGTGGACGAGTTGGCGCGGTTTCTTCTCCTCAAATATCA 3027

Qy        522 AACAAAAGAGCCTGTCACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGA 581
              |    ||  |||||  |||||||||||||||||||||||||  ||||||||  | |||  | ||
Db       3028 AGTGAAGCAGCCTATCACAAAGGCAGAGATGCTGACGAATGTCATCAGCAGGTACACGGG 3087

Qy        582 CTATTTTCCCATGATCTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGC 641
              |||  |||||   |||||||  ||||||||  ||||||||||||  ||  ||||||||||| |
```

```
Db       3088 CTACTTTCCTGTGATCTTCAGGAAAGCCCGTGAGTTCATAGAGATACTTTTTGGCATTTC 3147

Qy        642 CCTGACTGATATGGACCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCAC 701
              ||||| ||  |||||||  |   |  |||||||| ||||| | ||||||||||||||||
Db       3148 CCTGAGAGAAGTGGACCCTG---ATGACTCCTATGTCTTTGTAAACACATTAGACCTCAC 3204

Qy        702 CTATGAGGGAAGCCTGATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTAT 761
              || |||||  | |||| |||||| ||||||||| || ||||| | ||||||||||||||
Db       3205 CTCTGAGGGGTGTCTGAGTGATGAGCAGGGCATGTCCCAGAACCGCCTCCTGATTCTTAT 3264

Qy        762 TCTCAGTATGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTT 821
              ||| ||||| |||||||||||||||||| || || | | |||||||||||||||||| ||| |
Db       3265 TCTGAGTATCATCTTCATAAAGGGCACCTATGCCTCTGAGGAGGTCATCTGGGATGTGCT 3324

Qy        822 GAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCT 881
              ||||| |||||||||| |||||||||||||||||||||    | |||||| ||||  ||||
Db       3325 GAGTGGAATAGGGGTGCGTGCTGGGAGGGAGCACTTTGCCTTTGGGGAGCCCAGGGAGCT 3384

Qy        882 GCTCACTATACATTGGGTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGC 941
              |||||||  |  |||||||||  | | |||| |||||||||||||||||||||  | |
Db       3385 CCTCACTAAAGTTTGGGTGCAGGAACATTACCTAGAGTACCGGGAGGTGCCCAACTCTTC 3444

Qy        942 TCCTCCACGTTATGAATTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAG--A 999
              |||||| |||||  |||||  |||||||||||||||| |||||||  | ||  |||||  |
Db       3445 TCCTCCTCGTTACGAATTCCTGTGGGGTCCAAGAGCTCATTCAGAAGTCATTAAGAGGAA 3504

Qy       1000 AGTCTTAGAGTTTTTATCCAAGCTATCCAGTATCATCCCTA 1040
              |||  ||||||||||  ||| ||||  | || | ||||||
Db       3505 AGTAGTAGAGTTTTTGGCCATGCTAAAGAATACCGTCCCTA 3545
```

RESULT 7
US-08-845-528C-9
; Sequence 9, Application US/08845528C
; Patent No. 6027924
; GENERAL INFORMATION:
; APPLICANT: LUCAS, Sophie;
; APPLICANT: DE SMET, Charles;
; APPLICANT: BOON-FALLEUR, Thierry
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
; TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,528C
; FILING DATE: April 25, 1997
; CLASSIFICATION: 4335
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5455
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4225 base pairs
; TYPE: nucleic acids
; STRANDEDNESS: double-stranded
; TOPOLOGY: linear
US-08-845-528C-9

```
  Query Match           39.5%;  Score 411.4;  DB 3;  Length 4225;
  Best Local Similarity  74.1%;  Pred. No. 3.6e-97;
  Matches  608;  Conservative    0;  Mismatches  166;  Indels   47;  Gaps    5;

Qy        264 TCTTCCCCAGAGTCCTCCTGAGATTCC---TCCCCAGGGTCCTCCCAAGATCTCTCCCCA 320
              |||| |||||||| |||||| ||   |   |||        || |  ||   | || |
Db       2728 TCTTTCCCAGAGTTCTCCTGTGAGCTCCTTCCCCTCCTCCACTTCATCGAGTCTTTCCAA 2787

Qy        321 GGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCATCCCCTCTTTTGTG 380
              | || | ||  |||||||||| |||||||||  |    ||||| ||| ||| |  ||  |
Db       2788 GAGTTCCCCTGAGAGTCCTCTCCAGAGTCCTGTGATCTCCTTCTCCTCCTCCACTTCATT 2847

Qy        381 GACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAGGATACAGCTACTTGGCATGC 437
              || || |||   ||| ||||||||||     | ||| || ||||| | | |   | |
Db       2848 GAGCCCATTCAGTGAAGAGTCCAGCAGCCCAGTAGATGAATATACAAGTTCCTCAGACAC 2907

Qy        438 CTTGCCAGAAAGTGAATCCTTGCC---------------------------------- 461
              ||||| ||| |||||  |||||| |
Db       2908 CTTGCTAGAGAGTGATTCCTTGACAGACAGCGAGTCCTTGATAGAGAGCGAGCCCTTGTT 2967

Qy        462 CAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCA 521
              ||  ||| | |||||||||||||||  |||||| || ||||||||||||||||||||||
Db       2968 CACTTATACACTGGATGAAAAGGTGGACGAGTTGGCGCGGTTTCTTCTCCTCAAATATCA 3027

Qy        522 AACAAAAGAGCCTGTCACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGA 581
              |  ||  ||||| |||||||||||||||||||||||||| |||||||  | ||| | ||
Db       3028 AGTGAAGCAGCCTATCACAAAGGCAGAGATGCTGACGAATGTCATCAGCAGGTACACGGG 3087

Qy        582 CTATTTTCCCATGATCTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGC 641
              ||| |||||  |||||||| ||||||||| |||||||||||||| || ||||||||||| |
Db       3088 CTACTTTCCTGTGATCTTCAGGAAAGCCCGTGAGTTCATAGAGATACTTTTTGGCATTTC 3147

Qy        642 CCTGACTGATATGGACCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCAC 701
              |||||  ||  |||||||  |   |  |||||||| ||||| | ||||||||||||||||
Db       3148 CCTGAGAGAAGTGGACCCTG---ATGACTCCTATGTCTTTGTAAACACATTAGACCTCAC 3204

Qy        702 CTATGAGGGAAGCCTGATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTAT 761
              || ||||||  | |||| |||||| ||||||||| || ||||| | ||||||||||||||
Db       3205 CTCTGAGGGGTGTCTGAGTGATGAGCAGGGCATGTCCCAGAACCGCCTCCTGATTCTTAT 3264

Qy        762 TCTCAGTATGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTT 821
              ||| ||||| |||||||||||||||| || || | | |||||||||||||||||| ||| |
Db       3265 TCTGAGTATCATCTTCATAAAGGGCACCTATGCCTCTGAGGAGGTCATCTGGGATGTGCT 3324

Qy        822 GAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCT 881
              ||||| |||||||||| |||||||||||||||||||||    | |||||| ||||  ||||
Db       3325 GAGTGGAATAGGGGTGCGTGCTGGGAGGGAGCACTTTGCCTTTGGGGAGCCCAGGGAGCT 3384

Qy        882 GCTCACTATACATTGGGTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGC 941
              |||||||  |  |||||||||  | | |||| |||||||||||||||||||||  | |
Db       3385 CCTCACTAAAGTTTGGGTGCAGGAACATTACCTAGAGTACCGGGAGGTGCCCAACTCTTC 3444

Qy        942 TCCTCCACGTTATGAATTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAG--A 999
              |||||| |||||  |||||  |||||||||||||||| |||||||  | ||  |||||  |
Db       3445 TCCTCCTCGTTACGAATTCCTGTGGGGTCCAAGAGCTCATTCAGAAGTCATTAAGAGGAA 3504

Qy       1000 AGTCTTAGAGTTTTTATCCAAGCTATCCAGTATCATCCCTA 1040
              |||  ||||||||||  ||| ||||  | || | ||||||
Db       3505 AGTAGTAGAGTTTTTGGCCATGCTAAAGAATACCGTCCCTA 3545
```

RESULT 8
US-09-066-281B-9
; Sequence 9, Application US/09066281B
; Patent No. 6475783
; GENERAL INFORMATION:
; APPLICANT: LUCAS, Sophie; DE SMET, Charles; BOON-FALLEUR, Thierry
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING
; TITLE OF INVENTION: FOR TUMOR REJECTION ANTIGEN PRECURSOR MAGE-C1 AND MAGE-C2
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,281B
; FILING DATE: April 24, 1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/845,528
; FILING DATE: April 25, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5455.2 US - JEL/MAS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 318-3100
; TELEFAX: (212) 752-5958
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4225 base pairs
; TYPE: nucleic acids
; STRANDEDNESS: double-stranded
; TOPOLOGY: linear
US-09-066-281B-9

Query Match 39.5%; Score 411.4; DB 4; Length 4225;
Best Local Similarity 74.1%; Pred. No. 3.6e-97;
Matches 608; Conservative 0; Mismatches 166; Indels 47; Gaps 5;

```
Qy       264 TCTTCCCCAGAGTCCTCCTGAGATTCC---TCCCCAGGGTCCTCCCAAGATCTCTCCCCA 320
Db      2728 TCTTTCCCAGAGTTCTCCTGTGAGCTCCTTCCCCTCCTCCACTTCATCGAGTCTTTCCAA 2787

Qy       321 GGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCATCCCCTCTTTTGTG 380
Db      2788 GAGTTCCCCTGAGAGTCCTCTCCAGAGTCCTGTGATCTCCTTCTCCTCCTCCACTTCATT 2847

Qy       381 GACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAGGATACAGCTACTTGGCATGC 437
Db      2848 GAGCCCATTCAGTGAAGAGTCCAGCAGCCCAGTAGATGAATATACAAGTTCCTCAGACAC 2907

Qy       438 CTTGCCAGAAAGTGAATCCTTGCC---------------------------------- 461
Db      2908 CTTGCTAGAGAGTGATTCCTTGACAGACAGCGAGTCCTTGATAGAGAGCGAGCCCTTGTT 2967

Qy       462 CAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCA 521
Db      2968 CACTTATACACTGGATGAAAAGGTGGACGAGTTGGCGCGGTTTCTTCTCCTCAAATATCA 3027

Qy       522 AACAAAAGAGCCTGTCACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGA 581
Db      3028 AGTGAAGCAGCCTATCACAAAGGCAGAGATGCTGACGAATGTCATCAGCAGGTACACGGG 3087

Qy       582 CTATTTTCCCATGATCTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGC 641
Db      3088 CTACTTTCCTGTGATCTTCAGGAAAGCCCGTGAGTTCATAGAGATACTTTTTGGCATTTC 3147

Qy       642 CCTGACTGATATGGACCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCAC 701
Db      3148 CCTGAGAGAAGTGGACCCTG---ATGACTCCTATGTCTTTGTAAACACATTAGACCTCAC 3204

Qy       702 CTATGAGGGAAGCCTGATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTAT 761
Db      3205 CTCTGAGGGGTGTCTGAGTGATGAGCAGGGCATGTCCCAGAACCGCCTCCTGATTCTTAT 3264
```

```
Qy       762 TCTCAGTATGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTT 821
Db      3265 TCTGAGTATCATCTTCATAAAGGGCACCTATGCCTCTGAGGAGGTCATCTGGGATGTGCT 3324

Qy       822 GAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCT 881
Db      3325 GAGTGGAATAGGGGTGCGTGCTGGGAGGGAGCACTTTGCCTTTGGGGAGCCCAGGGAGCT 3384

Qy       882 GCTCACTATACATTGGGTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGC 941
Db      3385 CCTCACTAAAGTTTGGGTGCAGGAACATTACCTAGAGTACCGGGAGGTGCCCAACTCTTC 3444

Qy       942 TCCTCCACGTTATGAATTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAG--A 999
Db      3445 TCCTCCTCGTTACGAATTCCTGTGGGGTCCAAGAGCTCATTCAGAAGTCATTAAGAGGAA 3504

Qy      1000 AGTCTTAGAGTTTTTATCCAAGCTATCCAGTATCATCCCTA 1040
Db      3505 AGTAGTAGAGTTTTTGGCCATGCTAAAGAATACCGTCCCTA 3545
```

RESULT 9
US-09-061-709-1
; Sequence 1, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 1
; LENGTH: 4265
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-1

Query Match 39.5%; Score 411.4; DB 3; Length 4265;
Best Local Similarity 74.1%; Pred. No. 3.6e-97;
Matches 608; Conservative 0; Mismatches 166; Indels 47; Gaps 5;

```
Qy       264 TCTTCCCCAGAGTCCTCCTGAGATTCC---TCCCCAGGGTCCTCCCAAGATCTCTCCCCA 320
Db      2758 TCTTTCCCAGAGTTCTCCTGTGAGCTCCTTCCCCTCCTCCACTTCATCGAGTCTTTCCAA 2817

Qy       321 GGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCATCCCCTCTTTTGTG 380
Db      2818 GAGTTCCCCTGAGAGTCCTCTCCAGAGTCCTGTGATCTCCTTCTCCTCCTCCACTTCATT 2877

Qy       381 GACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAGGATACAGCTACTTGGCATGC 437
Db      2878 GAGCCCATTCAGTGAAGAGTCCAGCAGCCCAGTAGATGAATATACAAGTTCCTCAGACAC 2937

Qy       438 CTTGCCAGAAAGTGAATCCTTGCC---------------------------------- 461
Db      2938 CTTGCTAGAGAGTGATTCCTTGACAGACAGCGAGTCCTTGATAGAGAGCGAGCCCTTGTT 2997

Qy       462 CAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCA 521
Db      2998 CACTTATACACTGGATGAAAAGGTGGACGAGTTGGCGCGGTTTCTTCTCCTCAAATATCA 3057

Qy       522 AACAAAAGAGCCTGTCACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGA 581
Db      3058 AGTGAAGCAGCCTATCACAAAGGCAGAGATGCTGACGAATGTCATCAGCAGGTACACGGG 3117
```

```
Qy        582 CTATTTTCCCATGATCTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGC 641
              ||| |||||  |||||||| |||||||| |||||||||||| || ||||||||||| |
Db       3118 CTACTTTCCTGTGATCTTCAGGAAAGCCCGTGAGTTCATAGAGATACTTTTTGGCATTTC 3177

Qy        642 CCTGACTGATATGGACCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCAC 701
              |||||  ||  |||||||  |   |||||||| |||||| | ||||||||||||||||
Db       3178 CCTGAGAGAAGTGGACCCTG---ATGACTCCTATGTCTTTGTAAACACATTAGACCTCAC 3234

Qy        702 CTATGAGGGAAGCCTGATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTAT 761
              || ||||||  | |||| |||||| |||||||| || ||||| | ||||||||||||||
Db       3235 CTCTGAGGGGTGTCTGAGTGATGAGCAGGGCATGTCCCAGAACCGCCTCCTGATTCTTAT 3294

Qy        762 TCTCAGTATGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTT 821
              ||| ||||| |||||||||||||||| || || | | |||||||||||||||||| ||| |
Db       3295 TCTGAGTATCATCTTCATAAAGGGCACCTATGCCTCTGAGGAGGTCATCTGGGATGTGCT 3354

Qy        822 GAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCT 881
              ||||| |||||||||| |||||||||||||||||||    | |||||| |||||  ||||
Db       3355 GAGTGGAATAGGGGTGCGTGCTGGGAGGGAGCACTTTGCCTTTGGGGAGCCCAGGGAGCT 3414

Qy        882 GCTCACTATACATTGGGTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGC 941
              |||||||  |  |||||||||||  | | ||||| |||||||||||||||||||||  | |
Db       3415 CCTCACTAAAGTTTGGGTGCAGGAACATTACCTAGAGTACCGGGAGGTGCCCAACTCTTC 3474

Qy        942 TCCTCCACGTTATGAATTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAG--A 999
              |||||| ||||| |||||  |||||||||||||||| |||||||| | ||  |||||  |
Db       3475 TCCTCCTCGTTACGAATTCCTGTGGGGTCCAAGAGCTCATTCAGAAGTCATTAAGAGGAA 3534

Qy       1000 AGTCTTAGAGTTTTTATCCAAGCTATCCAGTATCATCCCTA 1040
              |||  |||||||||| ||| ||||  | || | ||||||
Db       3535 AGTAGTAGAGTTTTTGGCCATGCTAAAGAATACCGTCCCTA 3575
```

RESULT 10
US-09-899-651-1
; Sequence 1, Application US/09899651
; Patent No. 6576756
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
; TITLE OF INVENTION: Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/899,651
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/061,709
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 1
; LENGTH: 4265
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-899-651-1

Query Match 39.5%; Score 411.4; DB 4; Length 4265;
Best Local Similarity 74.1%; Pred. No. 3.6e-97;
Matches 608; Conservative 0; Mismatches 166; Indels 47; Gaps 5;

```
Qy        264 TCTTCCCCAGAGTCCTCCTGAGATTCC---TCCCCAGGGTCCTCCCAAGATCTCTCCCCA 320
              |||| |||||||| |||||| ||   |   |||       || |   ||   | || |
Db       2758 TCTTTCCCAGAGTTCTCCTGTGAGCTCCTTCCCCTCCTCCACTTCATCGAGTCTTTCCAA 2817

Qy        321 GGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCATCCCCTCTTTTGTG 380
              | || | ||  |||||||| |||||||||  |   ||||| ||| ||| |   ||  |
Db       2818 GAGTTCCCCTGAGAGTCCTCTCCAGAGTCCTGTGATCTCCTTCTCCTCCTCCACTTCATT 2877
```

```
Qy        381 GACCCGATTGGATGAGGAGTCCAGCAG---TGAAGAGGAGGATACAGCTACTTGGCATGC 437
              || || |||   ||| ||||||||||     | ||| || |||||  | | |   | |
Db       2878 GAGCCCATTCAGTGAAGAGTCCAGCAGCCCAGTAGATGAATATACAAGTTCCTCAGACAC 2937

Qy        438 CTTGCCAGAAAGTGAATCCTTGCC---------------------------------- 461
              ||||| ||| ||||| |||||| |
Db       2938 CTTGCTAGAGAGTGATTCCTTGACAGACAGCGAGTCCTTGATAGAGAGCGAGCCCTTGTT 2997

Qy        462 CAGGTATGCCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCA 521
              ||  ||| | |||||||||||||||||  |||||||  || ||||||||||||||||||
Db       2998 CACTTATACACTGGATGAAAAGGTGGACGAGTTGGCGCGGTTTCTTCTCCTCAAATATCA 3057

Qy        522 AACAAAAGAGCCTGTCACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGA 581
              |   ||  ||||| |||||||||||||||||||||||| |||||||  | ||| | ||
Db       3058 AGTGAAGCAGCCTATCACAAAGGCAGAGATGCTGACGAATGTCATCAGCAGGTACACGGG 3117

Qy        582 CTATTTTCCCATGATCTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGC 641
              ||| |||||  |||||||| |||||||| |||||||||||| || ||||||||||| |
Db       3118 CTACTTTCCTGTGATCTTCAGGAAAGCCCGTGAGTTCATAGAGATACTTTTTGGCATTTC 3177

Qy        642 CCTGACTGATATGGACCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCAC 701
              |||||  ||  |||||||  |   |||||||| |||||| | ||||||||||||||||
Db       3178 CCTGAGAGAAGTGGACCCTG---ATGACTCCTATGTCTTTGTAAACACATTAGACCTCAC 3234

Qy        702 CTATGAGGGAAGCCTGATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTAT 761
              || ||||||  | |||| |||||| |||||||| || ||||| | ||||||||||||||
Db       3235 CTCTGAGGGGTGTCTGAGTGATGAGCAGGGCATGTCCCAGAACCGCCTCCTGATTCTTAT 3294

Qy        762 TCTCAGTATGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTT 821
              ||| ||||| |||||||||||||||| || || | | |||||||||||||||||| ||| |
Db       3295 TCTGAGTATCATCTTCATAAAGGGCACCTATGCCTCTGAGGAGGTCATCTGGGATGTGCT 3354

Qy        822 GAGTGCAATAGGGGTGTGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCT 881
              ||||| |||||||||| |||||||||||||||||||    | |||||| |||||  ||||
Db       3355 GAGTGGAATAGGGGTGCGTGCTGGGAGGGAGCACTTTGCCTTTGGGGAGCCCAGGGAGCT 3414

Qy        882 GCTCACTATACATTGGGTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGC 941
              |||||||  |  |||||||||||  | | ||||| |||||||||||||||||||||  | |
Db       3415 CCTCACTAAAGTTTGGGTGCAGGAACATTACCTAGAGTACCGGGAGGTGCCCAACTCTTC 3474

Qy        942 TCCTCCACGTTATGAATTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAG--A 999
              |||||| ||||| |||||  |||||||||||||||| |||||||| | ||  |||||  |
Db       3475 TCCTCCTCGTTACGAATTCCTGTGGGGTCCAAGAGCTCATTCAGAAGTCATTAAGAGGAA 3534

Qy       1000 AGTCTTAGAGTTTTTATCCAAGCTATCCAGTATCATCCCTA 1040
              |||  |||||||||| ||| ||||  | || | ||||||
Db       3535 AGTAGTAGAGTTTTTGGCCATGCTAAAGAATACCGTCCCTA 3575
```

RESULT 11
US-08-724-774B-3
; Sequence 3, Application US/08724774B
; Patent No. 5908778
; GENERAL INFORMATION:
; APPLICANT: Rimoldi, Donata; Jongeneel, Victor; Coulie,
; APPLICANT: Pierre; Cerrottini, Jean-Charles; Carrel,
; APPLICANT: Stefan; Reed, Daryl
; TITLE OF INVENTION: MAGE-10 ENCODING cDNA, The Tumor
; TITLE OF INVENTION: Rejection Antigen Precuros Mage-10,
; TITLE OF INVENTION: Antibodies Specific To The Molecule, and
; TITLE OF INVENTION: Uses Thereof
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
; COMPUTER: IBM

```
;       OPERATING SYSTEM:  PC-DOS
;       SOFTWARE:  Wordperfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/724,774B
;       FILING DATE:  03-October-1996
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Hanson, No. 5908778man D.
;       REGISTRATION NUMBER:  30,946
;       REFERENCE/DOCKET NUMBER:  LUD 5457
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 688-9200
;       TELEFAX:  (212) 838-3884
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2559 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-724-774B-3

  Query Match             32.9%;  Score 342.2;  DB 2;  Length 2559;
  Best Local Similarity   69.7%;  Pred. No. 2.6e-79;
  Matches  498;  Conservative    0;  Mismatches  199;  Indels   18;  Gaps    2;

Qy        308 AGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCAT 367
              || | ||| | | | |  | |  | |||||||||||||| |||     |||||| |
Db        574 AGGTTTCTGCTGATGATGAGACACCAAATCCTCCCCAGAGTGCTCAGATAGCCTGCTCCT 633

Qy        368 CCC---------------CTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAG 412
              |||               | |||   |      | ||  |||||||  ||||||||  ||
Db        634 CCCCCTCGGTCGTTGCTTCCCTTCCATTAGATCAATCTGATGAGGGCTCCAGCAGCCAAA 693

Qy        413 AGGAGGA---TACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATG 469
              |||||||   | ||   ||    || | | |||||| |||| || || |||||   ||
Db        694 AGGAGGAGAGTCCAAGCACCCTACAGGTCCTGCCAGACAGTGAGTCTTTACCCAGAAGTG 753

Qy        470 CCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAG 529
                | |||||||||||| |||| |||||||||||||| ||| |||| |||||||   || |
Db        754 AGATAGATGAAAAGGTGACTGATTTGGTGCAGTTTCTGCTCTTCAAGTATCAAATGAAGG 813

Qy        530 AGCCTGTCACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTC 589
              ||||  |||||||||||||| || |||  || |||||| || || ||| | ||| | ||| |
Db        814 AGCCGATCACAAAGGCAGAAATACTGGAGAGTGTCATAAAAAATTATGAAGACCACTTCC 873

Qy        590 CCATGATCTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTG 649
              |  || | ||  |  |||||   |||| |||   |||  | ||||||||||   | | |
Db        874 CTTTGTTGTTTAGTGAAGCCTCCGAGTGCATGCTGCTGGTCTTTGGCATTGATGTAAAGG 933

Qy        650 ATATGGACCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGG 709
              |  |||| |||     |||||||| | || |||    | |  | | ||||||||||| |
Db        934 AAGTGGATCCCACTGGCCACTCCTTTGTCCTTGTCACCTCCCTGGGCCTCACCTATGATG 993

Qy        710 GAAGCCTGATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTA 769
              | | |||| |||||  |||| |||||||||||    |  | || || ||||| || || |
Db        994 GGATGCTGAGTGATGTCCAGAGCATGCCCAAGACTGGCATTCTCATACTTATCCTAAGCA 1053

Qy        770 TGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAA 829
              | |||||||||  |||||   |||   ||| ||||||||||||||||||    ||| | |
Db       1054 TAATCTTCATAGAGGGCTACTGCACCCCTGAGGAGGTCATCTGGGAAGCACTGAATATGA 1113

Qy        830 TAGGGGTGTGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTA 889
              | ||| ||| || ||||| |||||| | || |||||||| |||| ||||||||||
Db       1114 TGGGGCTGTATGATGGGATGGAGCACCTCATTTATGGGGAGCCCAGGAAGCTGCTCACCC 1173

Qy        890 TACATTGGGTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCAC 949
              | ||||||||||| ||| |||||||||||||| |||||||   ||||| |||| |||
Db       1174 AAGATTGGGTGCAGGAAAACTACCTGGAGTACCGGCAGGTGCCTGGCAGTGATCCTGCAC 1233

Qy        950 GTTATGAATTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAGAAGTCT 1004
              | ||||| ||| |||||||||||| || ||| | ||  || ||||  |||||
Db       1234 GGTATGAGTTTCTGTGGGGTCCAAGGGCTCATGCTGAAATTAGGAAGATGAGTCT 1288
```

```
RESULT 12
US-09-089-595-3
; Sequence 3, Application US/09089595
; Patent No. 6153728
;  GENERAL INFORMATION:
;    APPLICANT:  Rimoldi, Donata; Jongeneel, Victor; Coulie,
;    APPLICANT:  Pierre; Cerrottini, Jean-Charles; Carrel,
;    APPLICANT:  Stefan; Reed, Daryl
;    TITLE OF INVENTION:  MAGE-10 ENCODING cDNA, The Tumor
;    TITLE OF INVENTION:  Rejection Antigen Precuros Mage-10,
;    TITLE OF INVENTION:  Antibodies Specific To The Molecule, and
;    TITLE OF INVENTION:  Uses Thereof
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 144 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/089,595
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/724,774
;      FILING DATE:  03-October-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 6153728man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5457
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2559 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-089-595-3

  Query Match             32.9%;  Score 342.2;  DB 3;  Length 2559;
  Best Local Similarity   69.7%;  Pred. No. 2.6e-79;
  Matches  498;  Conservative    0;  Mismatches  199;  Indels   18;  Gaps    2;

Qy        308 AGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCAT 367
              || | ||| | | | |  | |  | |||||||||||||| |||     |||||| |
Db        574 AGGTTTCTGCTGATGATGAGACACCAAATCCTCCCCAGAGTGCTCAGATAGCCTGCTCCT 633

Qy        368 CCC---------------CTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAG 412
              |||               | |||   |      | ||  |||||||  ||||||||  ||
Db        634 CCCCCTCGGTCGTTGCTTCCCTTCCATTAGATCAATCTGATGAGGGCTCCAGCAGCCAAA 693

Qy        413 AGGAGGA---TACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATG 469
              |||||||   | ||   ||    || | | |||||| |||| || || |||||   ||
Db        694 AGGAGGAGAGTCCAAGCACCCTACAGGTCCTGCCAGACAGTGAGTCTTTACCCAGAAGTG 753

Qy        470 CCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAG 529
                | |||||||||||| |||| |||||||||||||| ||| |||| |||||||   || |
Db        754 AGATAGATGAAAAGGTGACTGATTTGGTGCAGTTTCTGCTCTTCAAGTATCAAATGAAGG 813

Qy        530 AGCCTGTCACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTC 589
              ||||  |||||||||||||| || |||  || |||||| || || ||| | ||| | ||| |
Db        814 AGCCGATCACAAAGGCAGAAATACTGGAGAGTGTCATAAAAAATTATGAAGACCACTTCC 873

Qy        590 CCATGATCTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTG 649
```

```
Db        874 CTTTGTTGTTTAGTGAAGCCTCCGAGTGCATGCTGCTGGTCTTTGGCATTGATGTAAAGG 933

Qy        650 ATATGGACCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGG 709

Db        934 AAGTGGATCCCACTGGCCACTCCTTTGTCCTTGTCACCTCCCTGGGCCTCACCTATGATG 993

Qy        710 GAAGCCTGATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTA 769

Db        994 GGATGCTGAGTGATGTCCAGAGCATGCCCAAGACTGGCATTCTCATACTTATCCTAAGCA 1053

Qy        770 TGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAA 829

Db       1054 TAATCTTCATAGAGGGCTACTGCACCCCTGAGGAGGTCATCTGGGAAGCACTGAATATGA 1113

Qy        830 TAGGGGTGTGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTA 889

Db       1114 TGGGGCTGTATGATGGGATGGAGCACCTCATTTATGGGGAGCCCAGGAAGCTGCTCACCC 1173

Qy        890 TACATTGGGTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCAC 949

Db       1174 AAGATTGGGTGCAGGAAAACTACCTGGAGTACCGGCAGGTGCCTGGCAGTGATCCTGCAC 1233

Qy        950 GTTATGAATTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAGAAGTCT 1004

Db       1234 GGTATGAGTTTCTGTGGGGTCCAAGGGCTCATGCTGAAATTAGGAAGATGAGTCT 1288
```

RESULT 13
US-09-382-855-3
; Sequence 3, Application US/09382855
; Patent No. 6174692
; GENERAL INFORMATION:
; APPLICANT: Rimoldi, Donata; Jongeneel, Victor; Coulie, Pierre;
; APPLICANT: Cerrottini, Jean-Charles; Carrel, Stefan; Reed, Daryl
; TITLE OF INVENTION: MAGE-10 ENCODING cDNA, The Tumor Rejection
; TITLE OF INVENTION: Antigen Precuros Mage-10, Antibodies Specific To The Molecule,
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski LLP
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/382,855
; FILING DATE: 25-August-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/089,595
; FILING DATE: 02-June-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,774
; FILING DATE: 03-October-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Schofield, Mary Anne
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5457.2 DIV - JEL/MAS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 318-3000
; TELEFAX: (212) 752-5958
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2559 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-09-382-855-3

Query Match 32.9%; Score 342.2; DB 3; Length 2559;
Best Local Similarity 69.7%; Pred. No. 2.6e-79;
Matches 498; Conservative 0; Mismatches 199; Indels 18; Gaps 2;

```
Qy        308 AGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCAT 367

Db        574 AGGTTTCTGCTGATGATGAGACACCAAATCCTCCCCAGAGTGCTCAGATAGCCTGCTCCT 633

Qy        368 CCC---------------CTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAG 412

Db        634 CCCCCTCGGTCGTTGCTTCCCTTCCATTAGATCAATCTGATGAGGGCTCCAGCAGCCAAA 693

Qy        413 AGGAGGA---TACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATG 469

Db        694 AGGAGGAGAGTCCAAGCACCCTACAGGTCCTGCCAGACAGTGAGTCTTTACCCAGAAGTG 753

Qy        470 CCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAG 529

Db        754 AGATAGATGAAAAGGTGACTGATTTGGTGCAGTTTCTGCTCTTCAAGTATCAAATGAAGG 813

Qy        530 AGCCTGTCACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTC 589

Db        814 AGCCGATCACAAAGGCAGAAATACTGGAGAGTGTCATAAAAAATTATGAAGACCACTTCC 873

Qy        590 CCATGATCTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTG 649

Db        874 CTTTGTTGTTTAGTGAAGCCTCCGAGTGCATGCTGCTGGTCTTTGGCATTGATGTAAAGG 933

Qy        650 ATATGGACCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGG 709

Db        934 AAGTGGATCCCACTGGCCACTCCTTTGTCCTTGTCACCTCCCTGGGCCTCACCTATGATG 993

Qy        710 GAAGCCTGATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTA 769

Db        994 GGATGCTGAGTGATGTCCAGAGCATGCCCAAGACTGGCATTCTCATACTTATCCTAAGCA 1053

Qy        770 TGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAA 829

Db       1054 TAATCTTCATAGAGGGCTACTGCACCCCTGAGGAGGTCATCTGGGAAGCACTGAATATGA 1113

Qy        830 TAGGGGTGTGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTA 889

Db       1114 TGGGGCTGTATGATGGGATGGAGCACCTCATTTATGGGGAGCCCAGGAAGCTGCTCACCC 1173

Qy        890 TACATTGGGTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCAC 949

Db       1174 AAGATTGGGTGCAGGAAAACTACCTGGAGTACCGGCAGGTGCCTGGCAGTGATCCTGCAC 1233

Qy        950 GTTATGAATTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAGAAGTCT 1004

Db       1234 GGTATGAGTTTCTGTGGGGTCCAAGGGCTCATGCTGAAATTAGGAAGATGAGTCT 1288
```

RESULT 14
US-09-183-714B-3
; Sequence 3, Application US/09183714B
; Patent No. 6221593
; GENERAL INFORMATION:
; APPLICANT: Boon-Falleur, Thierry
; APPLICANT: Brasseur, Francis
; APPLICANT: Rimoldi, Donata
; APPLICANT: De Plaen, Etienne
; TITLE OF INVENTION: Method for Determining Cancer by Determining Expression
; TITLE OF INVENTION: of MAGE-10
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/183,714B
; CURRENT FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: US 08/724,774
; PRIOR FILING DATE: 1996-10-03
; NUMBER OF SEQ ID NOS: 7
; SEQ ID NO 3

; LENGTH: 2559
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-183-714B-3

Query Match 32.9%; Score 342.2; DB 3; Length 2559;
Best Local Similarity 69.7%; Pred. No. 2.6e-79;
Matches 498; Conservative 0; Mismatches 199; Indels 18; Gaps 2;

```
Qy        308 AGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCAT 367
Db        574 AGGTTTCTGCTGATGATGAGACACCAAATCCTCCCCAGAGTGCTCAGATAGCCTGCTCCT 633

Qy        368 CCC---------------CTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAG 412
Db        634 CCCCCTCGGTCGTTGCTTCCCTTCCATTAGATCAATCTGATGAGGGCTCCAGCAGCCAAA 693

Qy        413 AGGAGGA---TACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATG 469
Db        694 AGGAGGAGAGTCCAAGCACCCTACAGGTCCTGCCAGACAGTGAGTCTTTACCCAGAAGTG 753

Qy        470 CCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAG 529
Db        754 AGATAGATGAAAAGGTGACTGATTTGGTGCAGTTTCTGCTCTTCAAGTATCAAATGAAGG 813

Qy        530 AGCCTGTCACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTC 589
Db        814 AGCCGATCACAAAGGCAGAAATACTGGAGAGTGTCATAAAAAATTATGAAGACCACTTCC 873

Qy        590 CCATGATCTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTG 649
Db        874 CTTTGTTGTTTAGTGAAGCCTCCGAGTGCATGCTGCTGGTCTTTGGCATTGATGTAAAGG 933

Qy        650 ATATGGACCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGG 709
Db        934 AAGTGGATCCCACTGGCCACTCCTTTGTCCTTGTCACCTCCCTGGGCCTCACCTATGATG 993

Qy        710 GAAGCCTGATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTA 769
Db        994 GGATGCTGAGTGATGTCCAGAGCATGCCCAAGACTGGCATTCTCATACTTATCCTAAGCA 1053

Qy        770 TGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAA 829
Db       1054 TAATCTTCATAGAGGGCTACTGCACCCCTGAGGAGGTCATCTGGGAAGCACTGAATATGA 1113

Qy        830 TAGGGGTGTGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTA 889
Db       1114 TGGGGCTGTATGATGGGATGGAGCACCTCATTTATGGGGAGCCCAGGAAGCTGCTCACCC 1173

Qy        890 TACATTGGGTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCAC 949
Db       1174 AAGATTGGGTGCAGGAAAACTACCTGGAGTACCGGCAGGTGCCTGGCAGTGATCCTGCAC 1233

Qy        950 GTTATGAATTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAGAAGTCT 1004
Db       1234 GGTATGAGTTTCTGTGGGGTCCAAGGGCTCATGCTGAAATTAGGAAGATGAGTCT 1288
```

RESULT 15
US-09-642-281-3
; Sequence 3, Application US/09642281
; Patent No. 6387698
; GENERAL INFORMATION:
; APPLICANT: Rimoldi, Donata; Jongeneel, Victor; Coulie, Pierre;
; Cerrottini, Jean-Charles; Carrel, Stefan; Reed, Daryl
; TITLE OF INVENTION: MAGE-10 ENCODING cDNA, The Tumor Rejection
; Antigen Precuros Mage-10, Antibodies Specific To The Molecu
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski LLP
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/642,281
; FILING DATE: 18-Aug-2000
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/089,595
; FILING DATE: 02-June-1998
; APPLICATION NUMBER: US 08/724,774
; FILING DATE: 03-October-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Schofield, Mary Anne
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5457.2 DIV - JEL/MAS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 318-3000
; TELEFAX: (212) 752-5958
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2559 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-642-281-3

Query Match 32.9%; Score 342.2; DB 4; Length 2559;
Best Local Similarity 69.7%; Pred. No. 2.6e-79;
Matches 498; Conservative 0; Mismatches 199; Indels 18; Gaps 2;

```
Qy        308 AGATCTCTCCCCAGGGTCCTCCGCAGAGTCCTCCCCAGAGTCCTCTAGACTCCTGCTCAT 367
Db        574 AGGTTTCTGCTGATGATGAGACACCAAATCCTCCCCAGAGTGCTCAGATAGCCTGCTCCT 633

Qy        368 CCC---------------CTCTTTTGTGGACCCGATTGGATGAGGAGTCCAGCAGTGAAG 412
Db        634 CCCCCTCGGTCGTTGCTTCCCTTCCATTAGATCAATCTGATGAGGGCTCCAGCAGCCAAA 693

Qy        413 AGGAGGA---TACAGCTACTTGGCATGCCTTGCCAGAAAGTGAATCCTTGCCCAGGTATG 469
Db        694 AGGAGGAGAGTCCAAGCACCCTACAGGTCCTGCCAGACAGTGAGTCTTTACCCAGAAGTG 753

Qy        470 CCCTGGATGAAAAGGTGGCTGAGTTGGTGCAGTTTCTTCTCCTCAAATATCAAACAAAAG 529
Db        754 AGATAGATGAAAAGGTGACTGATTTGGTGCAGTTTCTGCTCTTCAAGTATCAAATGAAGG 813

Qy        530 AGCCTGTCACAAAGGCAGAGATGCTGACGACTGTCATCAAGAAGTATAAGGACTATTTTC 589
Db        814 AGCCGATCACAAAGGCAGAAATACTGGAGAGTGTCATAAAAAATTATGAAGACCACTTCC 873

Qy        590 CCATGATCTTCGGGAAAGCCCATGAGTTCATAGAGCTAATTTTTGGCATTGCCCTGACTG 649
Db        874 CTTTGTTGTTTAGTGAAGCCTCCGAGTGCATGCTGCTGGTCTTTGGCATTGATGTAAAGG 933

Qy        650 ATATGGACCCCGACAACCACTCCTATTTCTTTGAAGACACATTAGACCTCACCTATGAGG 709
Db        934 AAGTGGATCCCACTGGCCACTCCTTTGTCCTTGTCACCTCCCTGGGCCTCACCTATGATG 993

Qy        710 GAAGCCTGATTGATGACCAGGGCATGCCCAAGAACTGTCTCCTGATTCTTATTCTCAGTA 769
Db        994 GGATGCTGAGTGATGTCCAGAGCATGCCCAAGACTGGCATTCTCATACTTATCCTAAGCA 1053

Qy        770 TGATCTTCATAAAGGGCAGCTGTGTCCCCGAGGAGGTCATCTGGGAAGTGTTGAGTGCAA 829
Db       1054 TAATCTTCATAGAGGGCTACTGCACCCCTGAGGAGGTCATCTGGGAAGCACTGAATATGA 1113
```

```
Qy        830 TAGGGGTGTGTGCTGGGAGGGAGCACTTTATATATGGGGATCCCAGAAAGCTGCTCACTA 889
              |  |||  |||  ||  |||||  |||||||  |  ||  ||||||||  |||||  ||||||||||
Db       1114 TGGGGCTGTATGATGGGATGGAGCACCTCATTTATGGGGAGCCCAGGAAGCTGCTCACCC 1173

Qy        890 TACATTGGGTGCAGAGAAAGTACCTGGAGTACCGGGAGGTGCCCAACAGTGCTCCTCCAC 949
              |  ||||||||||   |||  ||||||||||||||||  |||||||    |||||  ||||  |||
Db       1174 AAGATTGGGTGCAGGAAAACTACCTGGAGTACCGGCAGGTGCCTGGCAGTGATCCTGCAC 1233

Qy        950 GTTATGAATTTTTGTGGGGTCCAAGAGCCCATTCAGAGGCCAGCAAGAGAAGTCT 1004
              |  |||||  |||  ||||||||||||||  ||  |||  |  ||     ||  ||||   |||||
Db       1234 GGTATGAGTTTCTGTGGGGTCCAAGGGCTCATGCTGAAATTAGGAAGATGAGTCT 1288
```

Search completed: February 19, 2004, 20:01:01
Job time : 104 secs

XX
PR 01-AUG-1997; 97US-0905134.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153784/13.
DR P-PSDB; AAY12217.
XX
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from kidney, fetal kidney, dystrophic
PT muscle, muscle and heart tissue
XX
PS Claim 1; Page 389; 622pp; English.
XX
CC AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY01602 and
CC AAY11994 to AAY12260, respectively. The proteins given represent the
CC signal peptide and an N-terminal fragment of a secreted protein. The
CC nucleic acid sequences can be used for producing secreted human gene
CC products. They can also be used to develop products for diagnosis and
CC therapy. The proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used
CC for directing extracellular secretion of a polypeptide or the insertion
CC of a polypeptide into a membrane, or importing a polypeptide into
CC a cell.
XX
SQ Sequence 321 BP; 72 A; 96 C; 84 G; 68 T; 1 other;

Query Match 87.6%; Score 18.4; DB 20; Length 321;
Best Local Similarity 95.0%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGTCCTC 20
||||||||| ||||||||||
Db 107 TCATCCTCATCCTTGTCCTC 126

RESULT 15
AAH93187
ID AAH93187 standard; DNA; 700 BP.
XX
AC AAH93187;
XX
DT 09-OCT-2001 (first entry)
XX
DE Human inflammatory bowel disease related gene fragment IGR3480a.
XX
KW Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;
KW single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;
KW chromosome 5q31-33; forensic test; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200142511-A2.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US33632.
XX
PR 10-DEC-1999; 99US-0170257.
PR 10-APR-2000; 2000US-0196046.
XX
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
PA (ELLI-) ELLIPSIS BIOTHERAPEUTICS CORP.
XX
PI Daly M, Hudson TJ, Lander ES, Rioux J, Siminovitch K;
XX
DR WPI; 2001-367874/38.
XX
PT Testing for the presence of polymorphisms associated with inflammatory
PT bowel disease, using a hybridization assay -
XX
PS Disclosure; Page 431; 463pp; English.
XX
CC The present invention describes a method for detecting the presence of
CC polymorphisms associated with inflammatory bowel diseases such as
CC ulcerative colitis and Crohn's disease. The methods can be used to detect
CC the presence of genetic polymorphisms associated with inflammatory bowel
CC disease and correlating their occurrence with disease states. They may be
CC used in this way for phenotypic correlations, forensics, paternity
CC testing, medicine and genetic analysis. The present sequence is a gene
CC containing a polymorphic site described in the exemplification of the
CC invention.
XX
SQ Sequence 700 BP; 145 A; 174 C; 226 G; 151 T; 4 other;

Query Match 87.6%; Score 18.4; DB 22; Length 700;
Best Local Similarity 95.0%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CATCCTCACCCTTGTCCTCA 21
|| |||||||||||||||||
Db 420 CAGCCTCACCCTTGTCCTCA 439

Search completed: February 19, 2004, 21:28:32
Job time : 177.5 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 20:01:08 ; Search time 1596.5 Seconds
(without alignments)
538.116 Million cell updates/sec

Title: US-10-085-108-21_COPY_175_195
Perfect score: 21
Sequence: 1 TCATCCTCACCCTTGTCCTCA 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		21	100.0	124497	9	HS232G24	AL022152 Human DNA
2	c	20	95.2	1677	5	AF136525	AF136525 Gambusia
3		19.4	92.4	164601	2	AC125000	AC125000 Mus muscu
4	c	19.4	92.4	174412	9	AC009413	AC009413 Homo sapi
5	c	19.4	92.4	194991	2	AC122284	AC122284 Mus muscu
6	c	19	90.5	110807	10	AP003184	AP003184 Mus muscu
7		19	90.5	127623	2	AC133171	AC133171 Mus muscu
8	c	19	90.5	168135	2	AC120654	AC120654 Rattus no
9	c	19	90.5	263262	2	AC095135	AC095135 Rattus no
10	c	19	90.5	276523	10	AC012382	AC012382 Mus muscu
11		18.4	87.6	321	6	BD077670	BD077670 5'EST of
12	c	18.4	87.6	609	11	AU028771	AU028771 Rattus no
13		18.4	87.6	700	6	AX183446	AX183446 Sequence
14		18.4	87.6	993	8	VFABAMRCC	M16855 Broad bean
15		18.4	87.6	1227	9	AB020881	AB020881 Homo sapi
16		18.4	87.6	1246	6	E38387	E38387 Transcripti
17		18.4	87.6	1257	9	BC017559	BC017559 Homo sapi
18		18.4	87.6	1283	9	AF130312	AF130312 Homo sapi
19		18.4	87.6	1287	9	AF136570	AF136570 Homo sapi
20		18.4	87.6	1320	9	BC000381	BC000381 Homo sapi
21	c	18.4	87.6	2252	9	BC033499	BC033499 Homo sapi
22	c	18.4	87.6	2522	9	AK093365	AK093365 Homo sapi
23	c	18.4	87.6	36774	2	AC019572	AC019572 Drosophil
24	c	18.4	87.6	45005	9	AC004511	AC004511 Homo sapi
25	c	18.4	87.6	88442	10	AC005526	AC005526 Mus muscu
26	c	18.4	87.6	104824	9	HS73H22	AL035699 Human DNA
27	c	18.4	87.6	121493	10	AL645609	AL645609 Mouse DNA
28	c	18.4	87.6	150816	2	AC119532	AC119532 Rattus no
29	c	18.4	87.6	158387	2	AC095020	AC095020 Bos tauru
30	c	18.4	87.6	163569	2	AC023293	AC023293 Homo sapi
31	c	18.4	87.6	164946	9	AP000553	AP000553 Homo sapi
32		18.4	87.6	164974	3	AC011249	AC011249 Drosophil
33		18.4	87.6	166447	9	AC018751	AC018751 Homo sapi
34		18.4	87.6	168190	9	AC093841	AC093841 Homo sapi
35	c	18.4	87.6	169237	9	AC009516	AC009516 Homo sapi
36	c	18.4	87.6	169500	2	AC132291	AC132291 Mus muscu
37	c	18.4	87.6	170776	9	AL139087	AL139087 Human DNA
38	c	18.4	87.6	174376	3	AC092230	AC092230 Drosophil
39	c	18.4	87.6	174479	2	AC135613	AC135613 Pan trogl
40		18.4	87.6	178405	9	AC034216	AC034216 Homo sapi
41	c	18.4	87.6	180922	5	AL627168	AL627168 Zebrafish
42	c	18.4	87.6	183580	9	AL160253	AL160253 Human DNA
43		18.4	87.6	186135	2	AC129433	AC129433 Rattus no
44		18.4	87.6	186582	2	AC129249	AC129249 Rattus no
45		18.4	87.6	191704	2	AC139737	AC139737 Canis fam

ALIGNMENTS

RESULT 1
HS232G24
LOCUS HS232G24 124497 bp DNA linear PRI 09-MAR-2002
DEFINITION Human DNA sequence from clone RP6-232G24 on chromosome Xq27.1-27.3
Contains the gene for tha melanoma antigen gene family protein,
MAGEC3 and the MAGEC1 gene for melanoma antigen, family C,1,
complete sequence.
ACCESSION AL022152
VERSION AL022152.1 GI:3150086
KEYWORDS HTG; MAGEC1; MAGEC3.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 124497)

```
AUTHORS   Bird,C.
TITLE     Direct Submission
JOURNAL   Submitted (08-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
          Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
          humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On May 22, 1998 this sequence version replaced gi:2969932.
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations
          together with a note of the overlapping clone name. Note that the
          variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
          only a small overlap as described above.
          This sequence was finished as follows unless otherwise noted: all
          regions were either double-stranded or sequenced with an alternate
          chemistry or covered by high quality data (i.e., phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by at least
          one plasmid subclone or more than one M13 subclone; and the
          assembly was confirmed by restriction digest. The following
          abbreviations are used to associate primary accession numbers given
          in the feature table with their source databases: Em:, EMBL; Sw:,
          SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
          database can be found at
          http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
          was generated from part of bacterial clone contigs of human
          chromosome X, constructed by the Sanger Centre Chromosome X Mapping
          Group.  Further information can be found at
          http://www.sanger.ac.uk/HGP/ChrX
          RP6-232G24 is from the library RPCI-6 constructed by the group of
          Pieter de Jong. For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pPAC4
          This sequence is the entire insert of clone RP6-232G24 The true
          right end of clone RP3-326L12 is at 82755 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .124497
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP709G24232"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q27.1-27.3"
                     /clone="RP6-232G24"
                     /clone_lib="RPCI-6"
     repeat_region   6. .354
                     /note="L1MC2 repeat: matches 5195. .5546 of consensus"
     repeat_region   356. .762
                     /note="L1MC2 repeat: matches 5843. .6246 of consensus"
     repeat_region   763. .794
                     /note="16 copies 2 mer tg 93% conserved"
     repeat_region   946. .1161
                     /note="L1ME1 repeat: matches 5354. .5569 of consensus"
     repeat_region   1181. .1399
                     /note="L1ME1 repeat: matches 5914. .6136 of consensus"
     repeat_region   1471. .1606
                     /note="L2 repeat: matches 2574. .2708 of consensus"
     repeat_region   1939. .2330
                     /note="14 copies 28 mer 57% conserved"
     repeat_region   1969. .2326
                     /note="179 copies 2 mer gg 57% conserved"
     repeat_region   1972. .2331
                     /note="20 copies 18 mer 59% conserved"
     repeat_region   1996. .2265
                     /note="5 copies 54 mer 65% conserved"
     repeat_region   2029. .2181
                     /note="3 copies 51 mer 74% conserved"
     repeat_region   2153. .2347
                     /note="5 copies 39 mer 65% conserved"
     gene            2426. .4801
                     /gene="dA232G24.1"
     mRNA            join(2426. .2597,3291. .3384,3693. .4801)
                     /gene="dA232G24.1"
                     /product="dA232G24.1 (Melanoma antigen gene family
                     protein, MAGEC3)"
                     /note="(possible pseudogene)
                     match: cDNAs: Em:AF151378
                     match: ESTs: Em:BI460078"
                     /evidence=not_experimental
     repeat_region   2827. .2872
                     /note="23 copies 2 mer gg 78% conserved"
     CDS             3761. .4801
                     /gene="dA232G24.1"
                     /note="(possible pseudogene)
                     match: proteins: Sw:P43363 Tr:Q9BZ80 Tr:O60781 Tr:AAH04105
                     Tr:O95529 Tr:Q9R2A2 Sw:O15479 Tr:O89006 Tr:Q9BUN9
                     Tr:Q96M61"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="dA232G24.1 (Melanoma antigen gene family
                     protein, MAGEC3)"
                     /protein_id="CAA18146.1"
                     /db_xref="GI:3150087"
                     /translation="MPLFPNLPRLSFEEDFQNPSVTEDLVDAQDSIDEEEEDASSTSS
                     SSFHFLFPSSSSLSSSSPLSSPLPSTLILGVPEDEDMPAAGMPPLPQSPPEIPPQGPP
                     KISPQGPPQSPPQSPLDSCSSPLLWTRLDEESSSEEEDTATWHALPESESLPRYALDE
                     KVAELVQFLLLKYQTKEPVTKAEMLTTVIKKYKDYFPMIFGKAHEFIELIFGIALTDM
                     DPDNHSYFFEDTLDLTYEGSLIDDQGMPKNCLLILILSMIFIKGSCVPEEVIWEVLSA
                     IGVCAGREHFIYGDPRKLLTIHWVQRKYLEYREVPNSAPPRYEFLWGPRAHSEASKRS
                     LRVFIQAIQYHP"
     repeat_region   3876. .3959
                     /note="7 copies 12 mer 72% conserved"
     repeat_region   4027. .4110
                     /note="7 copies 12 mer 88% conserved"
     misc_feature    complement(5293. .5441)
                     /note="match: GSS: Em:AQ060261"
     repeat_region   6422. .6624
                     /note="MIR repeat: matches 47. .261 of consensus"
     repeat_region   8315. .8700
                     /note="MLT1B repeat: matches 1. .390 of consensus"
     repeat_region   9138. .9250
                     /note="MLT1B repeat: matches 1. .130 of consensus"
     repeat_region   9287. .9731
                     /note="L1MC3 repeat: matches 7166. .7618 of consensus"
     gene            11002. .16491
                     /gene="MAGEC1"
     mRNA            join(11002. .11086,11825. .11922,12078. .12184,12517. .16491)
                     /gene="MAGEC1"
                     /product="dA232G24.2 (Melanoma antigen gene family
                     protein, MAGEC1)"
                     /note="match: cDNAs: Em:AF056334
                     match: ESTs: Em:AI652057 Em:AL121366 Em:AI126114
                     Em:BG480822"
                     /evidence=not_experimental
     misc_feature    complement(11962. .12294)
                     /note="match: GSS: Em:AQ064665"
     CDS             join(12181. .12184,12517. .15941)
                     /gene="MAGEC1"
                     /note="match: proteins: Sw:O60732 Tr:O95529 Sw:Q9UBF1
                     Tr:Q96D45 Tr:Q99NC2 Tr:Q99PH7 Sw:Q9Y5V3"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="dA232G24.2 (Melanoma antigen gene family
                     protein, MAGEC1)"
                     /protein_id="CAD27434.1"
                     /db_xref="GI:20095260"
                     /translation="MGDKDMPTAGMPSLLQSSSESPQSCPEGEDSQSPLQIPQSSPES
                     DDTLYPLQSPQSRSEGEDSSDPLQRPPEGKDSQSPLQIPQSSPEGDDTQSPLQNSQSS
                     PEGKDSLSPLEISQSPPEGEDVQSPLQNPASSFFSSALLSIFQSSPESTQSPFEGFPQ
                     SVLQIPVSAASSSTLVSIFQSSPESTQSPFEGFPQSPLQIPVSRSFSSTLLSIFQSSP
                     ERTQSTFEGFAQSPLQIPVSPSSSSTLLSLFQSFSERTQSTFEGFAQSSLQIPVSPSF
                     SSTLVSLFQSSPERTQSTFEGFPQSPLQIPVSSSSSSTLLSLFQSSPERTHSTFEGFP
                     QSLLQIPMTSSFSSTLLSIFQSSPESAQSTFEGFPQSPLQIPGSPSFSSTLLSLFQSS
                     PERTHSTFEGFPQSPLQIPMTSSFSSTLLSILQSSPESAQSAFEGFPQSPLQIPVSSS
                     FSYTLLSLFQSSPERTHSTFEGFPQSPLQIPVSSSSSSSTLLSLFQSSPECTQSTFEG
                     FPQSPLQIPQSPPEGENTHSPLQIVPSLPEWEDSLSPHYFPQSPPQGEDSLSPHYFPQ
                     SPPQGEDSLSPHYFPQSPQGEDSLSPHYFPQSPPQGEDSMSPLYFPQSPLQGEEFQSS
```

```
                     LQSPVSICSSSTPSSLPQSFPESSQSPPEGPVQSPLHSPQSPPEGMHSQSPLQSPESA
                     PEGEDSLSPLQIPQSPLEGEDSLSSLHFPQSPPEWEDSLSPLHFPQFPPQGEDFQSSL
                     QSPVSICSSSTSLSLPQSFPESPQSPPEGPAQSPLQRPVSSFFSYTLASLLQSSHESP
                     QSPPEGPAQSPLQSPVSSFPSSTSSSLSQSSPVSSFPSSTSSSLSKSSPESPLQSPVI
                     SFSSSTSLSPFSEESSSPVDEYTSSSDTLLESDSLTDSESLIESEPLFTYTLDEKVDE
                     LARFLLLKYQVKQPITKAEMLTNVISRYTGYFPVIFRKAREFIEILFGISLREVDPDD
                     SYVFVNTLDLTSEGCLSDEQGMSQNRLLILILSIIFIKGTYASEEVIWDVLSGIGVRA
                     GREHFAFGEPRELLTKVWVQEHYLEYREVPNSSPPRYEFLWGPRAHSEVIKRKVVEFL
                     AMLKNTVPITFPSSYKDALKDVEERAQAIIDTTDDSTATESASSSVMSPSFSSE"
     repeat_region   12417. .12500
                     /note="7 copies 12 mer 70% conserved"
     repeat_region   12420. .12497
                     /note="13 copies 6 mer tcctcc 67% conserved"
     repeat_region   12675. .12752
                     /note="2 copies 39 mer 89% conserved"
     repeat_region   12721. .12912
                     /note="4 copies 48 mer 88% conserved"
     repeat_region   12928. .13977
                     /note="10 copies 105 mer 94% conserved"
     repeat_region   13844. .14719
                     /note="2 copies 438 mer 82% conserved"
     repeat_region   14023. .14262
                     /note="5 copies 48 mer 84% conserved"
     repeat_region   14039. .14266
                     /note="19 copies 12 mer 58% conserved"
     repeat_region   14217. .14306
                     /note="2 copies 45 mer 100% conserved"
     repeat_region   14455. .14742
                     /note="24 copies 12 mer 56% conserved"
     repeat_region   14480. .14719
                     /note="5 copies 48 mer 89% conserved"
     repeat_region   14683. .14997
                     /note="3 copies 105 mer 86% conserved"
     repeat_region   14948. .15049
                     /note="2 copies 51 mer 99% conserved"
     repeat_region   15169. .15222
                     /note="3 copies 18 mer 88% conserved"
     repeat_region   17466. .17718
                     /note="MIR repeat: matches 11. .261 of consensus"
     repeat_region   19287. .19670
                     /note="MLT1B repeat: matches 1. .390 of consensus"

  Query Match             100.0%;  Score 21;  DB 9;  Length 124497;
  Best Local Similarity   100.0%;  Pred. No. 67;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              |||||||||||||||||||||
Db       3935 TCATCCTCACCCTTGTCCTCA 3955


RESULT 2
AF136525/c
LOCUS       AF136525                1677 bp    DNA     linear   VRT 09-JUN-1999
DEFINITION  Gambusia affinis RAPD band OPD2sub(1560) sequence.
ACCESSION   AF136525
VERSION     AF136525.1  GI:5019775
KEYWORDS    .
SOURCE      Gambusia affinis (western mosquitofish)
  ORGANISM  Gambusia affinis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Cyprinodontiformes; Poeciliidae; Gambusia.
REFERENCE   1  (bases 1 to 1677)
  AUTHORS   Theodorakis,C.W. and Bickham,J.W.
  TITLE     Molecular characterization of contaminant-indicative RAPD markers:
            homology between species, DNA sequence, and development of
            marker-specific primers
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1677)
  AUTHORS   Theodorakis,C.W. and Bickham,J.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-1999) Wildlife and Fisheries Sciences, Texas A&M
            University, 210 Nagle Hall, College Station, TX 77843, USA
FEATURES             Location/Qualifiers
     source          1. .1677
                     /organism="Gambusia affinis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:33528"
     misc_feature    1. .1677
                     /note="RAPD band OPD2sub(1560)"
BASE COUNT      559 a    269 c    335 g    512 t      2 others
ORIGIN

  Query Match             95.2%;  Score 20;  DB 5;  Length 1677;
  Best Local Similarity   100.0%;  Pred. No. 3.4e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              ||||||||||||||||||||
Db       1518 TCATCCTCACCCTTGTCCTC 1499


RESULT 3
AC125000
LOCUS       AC125000               164601 bp    DNA     linear   HTG 27-MAR-2003
DEFINITION  Mus musculus clone RP24-409L6, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   AC125000
VERSION     AC125000.3  GI:29294292
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 164601)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-409L6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164601)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 164601)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
```

```
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 27, 2003 this sequence version replaced gi:28191558.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L25636
                Center clone name: 409_L_6
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 162739 bases at least Q40
                Consensus quality: 163377 bases at least Q30
                Consensus quality: 163717 bases at least Q20
                Insert size: 153000; agarose-fp
                Insert size: 163901; sum-of-contigs
                Quality coverage: 11.4 in Q20 bases; agarose-fp
                Quality coverage: 10.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     5870: contig of 5870 bp in length
            *    5871     5970: gap of 100 bp
            *    5971     9585: contig of 3615 bp in length
            *    9586     9685: gap of 100 bp
            *    9686    13730: contig of 4045 bp in length
            *   13731    13830: gap of 100 bp
            *   13831    32078: contig of 18248 bp in length
            *   32079    32178: gap of 100 bp
            *   32179    78827: contig of 46649 bp in length
            *   78828    78927: gap of 100 bp
            *   78928   106867: contig of 27940 bp in length
            *  106868   106967: gap of 100 bp
            *  106968   136363: contig of 29396 bp in length
            *  136364   136463: gap of 100 bp
            *  136464   164601: contig of 28138 bp in length.
FEATURES            Location/Qualifiers
     source         1. .164601
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10090"
                    /clone="RP24-409L6"
                    /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature   1. .5870
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:left"
     misc_feature   5971. .9585
                    /note="assembly_fragment"
     misc_feature   9686. .13730
                    /note="assembly_fragment"
     misc_feature   13831. .32078
                    /note="assembly_fragment"
     misc_feature   32179. .78827
                    /note="assembly_fragment"
     misc_feature   78928. .106867
                    /note="assembly_fragment"
     misc_feature   106968. .136363
                    /note="assembly_fragment"
     misc_feature   136464. .164601
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:right"
BASE COUNT    45950 a  37019 c  36895 g  44037 t    700 others
ORIGIN

  Query Match            92.4%;  Score 19.4;  DB 2;  Length 164601;
  Best Local Similarity  95.2%;  Pred. No. 3e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              || |||||||||||||||||||
Db       2662 TCCTCCTCACCCTTGTCCTCA 2682


RESULT 4
AC009413/c
LOCUS       AC009413              174412 bp    DNA     linear   PRI 07-OCT-2000
DEFINITION  Homo sapiens BAC clone RP11-459K11 from 2, complete sequence.
ACCESSION   AC009413
VERSION     AC009413.2  GI:10716654
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174412)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 174412)
  AUTHORS   Le,T., Drone,K. and Du,F.
  TITLE     The sequence of Homo sapiens BAC clone RP11-459K11
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 174412)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 174412)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Oct 7, 2000 this sequence version replaced gi:5757519.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0459K11
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.
```

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-478M12, 200 base pair overlap. Actual start of this clone is at base position 1 of RP11-459K11; actual end is at base position 174218 of RP11-459K11.

FEATURES Location/Qualifiers

source	1. .174412
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-459K11"
	/clone_lib="RPCI-11"
repeat_region	81. .885
	/rpt_family="MER21-group"
repeat_region	895. .1334
	/rpt_family="MaLR"
repeat_region	1491. .1807
	/rpt_family="Alu"
repeat_region	1973. .2174
	/rpt_family="Other"
repeat_region	2175. .2481
	/rpt_family="Alu"
repeat_region	2482. .2798
	/rpt_family="Other"
repeat_region	2799. .3000
	/rpt_family="L1"
repeat_region	3242. .3588
	/rpt_family="Retroviral"
repeat_region	3631. .3671
	/rpt_family="Retroviral"
repeat_region	3859. .4163
	/rpt_family="Alu"
repeat_region	5091. .5153
	/rpt_family="L2"
misc_feature	5123. .5137
	/note="similar to EST AA156077 (NID:g1727892) zo48g03.r1"
repeat_region	5156. .5212
	/rpt_family="MIR"
repeat_region	5352. .5651
	/rpt_family="Alu"
repeat_region	5713. .5956
	/rpt_family="MIR"
repeat_region	6019. .6058
	/rpt_family="Alu"
repeat_region	6059. .6212
	/rpt_family="Alu"
repeat_region	6258. .6507
	/rpt_family="L2"
repeat_region	6545. .6760
	/rpt_family="MER1_type"
repeat_region	8025. .8214
	/rpt_family="MIR"
repeat_region	8452. .8777
	/rpt_family="Alu"
misc_feature	9620. .9624
	/note="similar to EST D51468 (NID:g951704)"
repeat_region	9872. .10079
	/rpt_family="MER1_type"
repeat_region	10147. .10184
	/rpt_family="MIR"
misc_feature	10310. .10313
	/note="match to EST AA854979 (NID:g2942517) aj53f09.s1"
repeat_region	10652. .10743
	/rpt_family="MER53"
repeat_region	11256. .11416
	/rpt_family="L2"
misc_feature	12451. .12460
	/note="match to EST AA453518 (NID:g2167187) zx47f12.s1"
misc_feature	13521. .13731
	/note="match to EST AA854979 (NID:g2942517) aj53f09.s1"
misc_feature	13522. .13830
	/note="similar to EST AW183295 (NID:g6451893) xj75e08.x1"
misc_feature	13522. .13827
	/note="similar to EST AW118588 (NID:g6087172) xd94c07.x1"
misc_feature	13525. .13982
	/note="match to EST AI028214 (NID:g3245523) ov96c05.x1"
misc_feature	13525. .13827
	/note="match to EST AA453518 (NID:g2167187) zx47f12.s1"
misc_feature	13724. .13827
	/note="match to EST AA694503 (NID:g2695441) ah34h09.s1"
misc_feature	14176. .14403
	/note="similar to EST AW118588 (NID:g6087172) xd94c07.x1"
misc_feature	14176. .14299
	/note="match to EST AA694503 (NID:g2695441) ah34h09.s1"
misc_feature	14176. .14231
	/note="match to EST AA453518 (NID:g2167187) zx47f12.s1"
misc_feature	14264. .14665
	/note="match to EST AA453517 (NID:g2167186) zx47f12.r1"
repeat_region	14531. .14600
	/rpt_family="MIR"
misc_feature	14953. .14958
	/note="similar to EST AA488424 (NID:g2215855) ab39e04.s1"
repeat_region	15023. .15349
	/rpt_family="MER1_type"
repeat_region	15382. .15463
	/rpt_family="MIR"
repeat_region	15487. .15795
	/rpt_family="Alu"
repeat_region	16227. .16529
	/rpt_family="Alu"
repeat_region	16821. .16854
	/rpt_family="L2"
repeat_region	16832. .16980
	/rpt_family="MIR"
repeat_region	17090. .17258
	/rpt_family="MIR"
repeat_region	17312. .17410
	/rpt_family="L2"
repeat_region	18873. .19154
	/rpt_family="Alu"
repeat_region	19281. .19571
	/rpt_family="Alu"
repeat_region	20138. .20206
	/rpt_family="U4"
repeat_region	21599. .21725
	/rpt_family="L2"
repeat_region	21982. .22286
	/rpt_family="Alu"
misc_feature	23682. .23686
	/note="similar to EST AA888358 (NID:g3004033) nw79a02.s1"

```
     misc_feature    24110. .24377
                     /note="match to EST AA663026 (NID:g2617017) ab72a10.s1"
     repeat_region   24720. .25073
                     /rpt_family="MaLR"
     repeat_region   26879. .27174
                     /rpt_family="Alu"
     repeat_region   27212. .27516
                     /rpt_family="Alu"
     repeat_region   27555. .27718
                     /rpt_family="MIR"

  Query Match           92.4%;  Score 19.4;  DB 9;  Length 174412;
  Best Local Similarity   95.2%;  Pred. No. 3e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              |||||||||||||||||| ||
Db     121530 TCATCCTCACCCTTGTCCCCA 121510


RESULT 5
AC122284/c
LOCUS       AC122284               194991 bp    DNA     linear   HTG 23-MAY-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-246A5, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC122284
VERSION     AC122284.1  GI:21105140
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 194991)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 194991)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            ----------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: M_BA0246A05
            ------------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 190912 bases at least Q40
            Consensus quality: 192510 bases at least Q30
            Consensus quality: 193557 bases at least Q20
            Insert size: 204000; agarose-fp
            Insert size: 195984; sum-of-contigs
            Quality coverage: 11.70 in Q20 bases; agarose-fp
            Quality coverage: 6.63 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1173: contig of 1173 bp in length
            *     1174     1273: gap of unknown length
            *     1274     2330: contig of 1057 bp in length
            *     2331     2430: gap of unknown length
            *     2431     3593: contig of 1163 bp in length
            *     3594     3693: gap of unknown length
            *     3694     4999: contig of 1306 bp in length
            *     5000     5099: gap of unknown length
            *     5100     6465: contig of 1366 bp in length
            *     6466     6565: gap of unknown length
            *     6566     9541: contig of 2976 bp in length
            *     9542     9641: gap of unknown length
            *     9642    13055: contig of 3414 bp in length
            *    13056    13155: gap of unknown length
            *    13156    20585: contig of 7430 bp in length
            *    20586    20685: gap of unknown length
            *    20686    29723: contig of 9038 bp in length
            *    29724    29823: gap of unknown length
            *    29824    43883: contig of 14060 bp in length
            *    43884    43983: gap of unknown length
            *    43984    56074: contig of 12091 bp in length
            *    56075    56174: gap of unknown length
            *    56175    80163: contig of 23989 bp in length
            *    80164    80263: gap of unknown length
            *    80264   135342: contig of 55079 bp in length
            *   135343   135442: gap of unknown length
            *   135443   194991: contig of 59549 bp in length.
FEATURES             Location/Qualifiers
     source          1. .194991
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-246A5"
     misc_feature    1. .1173
                     /note="assembly_name:Contig9"
     misc_feature    1274. .2330
                     /note="assembly_name:Contig47"
     misc_feature    2431. .3593
                     /note="assembly_name:Contig50"
     misc_feature    3694. .4999
                     /note="assembly_name:Contig53"
     misc_feature    5100. .6465
                     /note="assembly_name:Contig57"
     misc_feature    6566. .9541
                     /note="assembly_name:Contig58"
     misc_feature    9642. .13055
                     /note="assembly_name:Contig59
                     clone_end:SP6
                     vector_side:left"
     misc_feature    13156. .20585
                     /note="assembly_name:Contig60"
     misc_feature    20686. .29723
                     /note="assembly_name:Contig61"
     misc_feature    29824. .43883
                     /note="assembly_name:Contig62"
     misc_feature    43984. .56074
                     /note="assembly_name:Contig63"
     misc_feature    56175. .80163
                     /note="assembly_name:Contig64
                     clone_end:T7
                     vector_side:left"
     misc_feature    80264. .135342
                     /note="assembly_name:Contig65"
     misc_feature    135443. .194991
                     /note="assembly_name:Contig66"
BASE COUNT    55728 a  42930 c  41683 g  53332 t   1318 others
ORIGIN

  Query Match           92.4%;  Score 19.4;  DB 2;  Length 194991;
  Best Local Similarity   95.2%;  Pred. No. 3e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
```

```
                   || ||||||||||||||||||
Db      83556 TCCTCCTCACCCTTGTCCTCA 83536


RESULT 6
AP003184/c
LOCUS       AP003184               110807 bp    DNA     linear   ROD 20-OCT-2001
DEFINITION  Mus musculus genomic DNA, chromosome 7 clone:B271I8, complete
            sequences.
ACCESSION   AP003184
VERSION     AP003184.2  GI:16303287
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Hattori,M., Kato,R., Toyoda,A., Taylor,T.D., Fujiyama,A., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     House mouse genomic DNA
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 110807)
  AUTHORS   Hattori,M., Kato,R., Toyoda,A., Taylor,T.D., Fujiyama,A., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2001) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Oct 19, 2001 this sequence version replaced gi:15208269.
            This work was done in collaboration with Suda, C., Shirohzu, H. and
            Sasaki, H.
            National Institute of Genetics
            Human Genetics Division
            1-111, Yata, Mishima, Shizuoka, JAPAN
            zip: 411-8540
            e-mail: hisasaki@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .110807
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="7"
                     /clone="B271I8"
BASE COUNT    29197 a  25790 c  27154 g  28666 t
ORIGIN

  Query Match             90.5%;  Score 19;  DB 10;  Length 110807;
  Best Local Similarity   100.0%;  Pred. No. 4.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CATCCTCACCCTTGTCCTC 20
              ||||||||||||||||||
Db      77186 CATCCTCACCCTTGTCCTC 77168


RESULT 7
AC133171
LOCUS       AC133171               127623 bp    DNA     linear   HTG 28-MAY-2003
DEFINITION  Mus musculus chromosome UNK clone RP24-271D14, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC133171
VERSION     AC133171.4  GI:31088442
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 127623)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 127623)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 127623)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On May 28, 2003 this sequence version replaced gi:22830432.

            ---------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information ------------------
            Center project name: M_BB0271D14
            -----------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 126778 bases at least Q40
            Consensus quality: 126819 bases at least Q30
            Consensus quality: 126834 bases at least Q20
            Insert size: 134000; agarose-fp
            Insert size: 127523; sum-of-contigs
            Quality coverage: 13.21 in Q20 bases; agarose-fp
            Quality coverage: 12.81 in Q20 bases; sum-of-contigs
            -----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    56481: contig of 56481 bp in length
            *    56482    56581: gap of unknown length
            *    56582   127623: contig of 71042 bp in length.
FEATURES             Location/Qualifiers
     source          1. .127623
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP24-271D14"
     misc_feature    1. .56481
                     /note="assembly_name:Contig22"
     misc_feature    56582. .127623
                     /note="assembly_name:Contig23"
BASE COUNT    34958 a  24419 c  25165 g  42981 t    100 others
ORIGIN

  Query Match             90.5%;  Score 19;  DB 2;  Length 127623;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCT 19
              |||||||||||||||||||
Db      54485 TCATCCTCACCCTTGTCCT 54503


RESULT 8
AC120654/c
LOCUS       AC120654               168135 bp    DNA     linear   HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-389M12, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
```

```
ACCESSION   AC120654
VERSION     AC120654.4  GI:25008188
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 168135)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168135)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 168135)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15, 2002 this sequence version replaced gi:23195506.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GXHU
                Center clone name: CH230-389M12
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 153795 bases at least Q40
                Consensus quality: 155395 bases at least Q30
                Consensus quality: 156543 bases at least Q20
                Estimated insert size: 154783; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7528: contig of 7528 bp in length
            *     7529     7628: gap of unknown length
            *     7629   165839: contig of 158211 bp in length
            *   165840   165939: gap of unknown length
            *   165940   168135: contig of 2196 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168135
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-389M12"
     misc_feature    1. .1195
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    2278. .4596
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    4496. .5381
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ184583"
     misc_feature    7629. .9074
                     /note="wgs_contig"
     misc_feature    161618. .163039
                     /note="wgs_contig"
     misc_feature    164009. .165839
                     /note="wgs_contig"
BASE COUNT    43739 a  34745 c  37114 g  41570 t  10967 others
ORIGIN

  Query Match             90.5%;  Score 19;  DB 2;  Length 168135;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          2 CATCCTCACCCTTGTCCTC 20
              ||||||||||||||||||||
Db      67161 CATCCTCACCCTTGTCCTC 67143


RESULT 9
AC095135/c
LOCUS       AC095135               263262 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-8C1, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC095135
VERSION     AC095135.10  GI:30467351
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 263262)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 263262)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 263262)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:24940774.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GCRK
                Center clone name: CH230-8C1
            -------------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 240713 bases at least Q40
                Consensus quality: 241900 bases at least Q30
                Consensus quality: 242534 bases at least Q20
                Estimated insert size: 261225; sum-of-contigs estimation
                Quality coverage: 10x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    17247: contig of 17247 bp in length
            *    17248    17347: gap of unknown length
            *    17348    20718: contig of 3371 bp in length
            *    20719    20818: gap of unknown length
            *    20819   260812: contig of 239994 bp in length
            *   260813   260912: gap of unknown length
            *   260913   263262: contig of 2350 bp in length.
FEATURES             Location/Qualifiers
     source          1. .263262
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-8C1"
BASE COUNT    60249 a  58455 c  60485 g  63790 t  20283 others
ORIGIN

  Query Match             90.5%;  Score 19;  DB 2;  Length 263262;
  Best Local Similarity   100.0%;  Pred. No. 4.2e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CATCCTCACCCTTGTCCTC 20
              ||||||||||||||||||||
Db     224415 CATCCTCACCCTTGTCCTC 224397
```

```
RESULT 10
AC012382/c
LOCUS       AC012382               276523 bp    DNA     linear   ROD 10-FEB-2001
DEFINITION  Mus musculus chromosome 7, clone RP23-92L23, complete sequence.
ACCESSION   AC012382
VERSION     AC012382.14  GI:12740210
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 276523)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 7, clone RP23-92L23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 276523)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 276523)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 10, 2001 this sequence version replaced gi:10280830.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2124
                Center clone name: 92_L_23
            --------------.
FEATURES             Location/Qualifiers
     source          1. .276523
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="7"
                     /map="7"
                     /clone="RP23-92L23"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     repeat_region   9. .58
                     /rpt_family="B1_MM"
     repeat_region   complement(82. .1042)
                     /rpt_family="Lx9"
     repeat_region   complement(1633. .1847)
                     /rpt_family="B3"
     repeat_region   complement(2088. .2150)
                     /rpt_family="MLT2B2"
     repeat_region   2161. .2262
                     /rpt_family="(TAGA)n"
     repeat_region   2295. .2441
                     /rpt_family="B1_MM"
     repeat_region   2444. .2480
                     /rpt_family="(A)n"
     repeat_region   2490. .2648
                     /rpt_family="(TAGA)n"
     repeat_region   3099. .3296
                     /rpt_family="MTD"
     repeat_region   3298. .3587
                     /rpt_family="Lx3"
     repeat_region   3605. .3653
                     /rpt_family="B1_MM"
     repeat_region   3839. .3985
                     /rpt_family="B1_MM"
     repeat_region   4326. .4537
                     /rpt_family="ORR1B"
     repeat_region   complement(5226. .5419)
                     /rpt_family="ORR1D"
     repeat_region   5420. .5568
                     /rpt_family="(GGGAGA)n"
     unsure          5489. .5504
                     /note="<30 qual SNGL region"
     unsure          5503
                     /note="Unresolved tandem repeat; possibly missing 50 bp
                     poly [GA]"
     repeat_region   complement(5569. .5668)
                     /rpt_family="ORR1D"
     repeat_region   6031. .6199
                     /rpt_family="B3"
     repeat_region   complement(6257. .6325)
                     /rpt_family="PB1D7"
     repeat_region   complement(9189. .9528)
                     /rpt_family="L1MB6"
     unsure          9584. .9596
                     /note="<30 qual SNGL region"
     repeat_region   9591. .9642
                     /rpt_family="(CCA)n"
     unsure          9644. .9694
                     /note="<30 qual SNGL region"
     unsure          9667
                     /note="probably C, possibly A"
     unsure          9729. .9730
                     /note="<30 qual SNGL region"
     repeat_region   complement(9898. .10110)
                     /rpt_family="B3"
     repeat_region   complement(11825. .12552)
                     /rpt_family="RMER19"
     repeat_region   15030. .15101
                     /rpt_family="(CTG)n"
     repeat_region   16328. .16364
                     /rpt_family="(GAAAA)n"
     repeat_region   17539. .17797
                     /rpt_family="(TCCA)n"
     repeat_region   17839. .17893
                     /rpt_family="(TTTC)n"
```

```
     repeat_region   complement(17894. .18386)
                     /rpt_family="L1_MM"
     repeat_region   18393. .18977
                     /rpt_family="RMER19A"
     repeat_region   19042. .19337
                     /rpt_family="(TCCA)n"
     repeat_region   22078. .22256
                     /rpt_family="(CAGAGA)n"
     repeat_region   22630. .23038
                     /rpt_family="Lx4"
     repeat_region   23040. .23073
                     /rpt_family="(CAAAA)n"
     repeat_region   23338. .23357
                     /rpt_family="(TG)n"
     repeat_region   25017. .25075
                     /rpt_family="B4A"
     repeat_region   25089. .25140
                     /rpt_family="(TG)n"
     repeat_region   25815. .26000
                     /rpt_family="CT-rich"
     repeat_region   25995. .26125
                     /rpt_family="(TCCA)n"
     repeat_region   26037. .26399
                     /rpt_family="CT-rich"
     repeat_region   26414. .26545
                     /rpt_family="(TCCA)n"
     repeat_region   26552. .26699
                     /rpt_family="CT-rich"
     repeat_region   complement(27906. .28013)
                     /rpt_family="B4A"
     repeat_region   complement(28114. .28643)
                     /rpt_family="L1MB3"
     repeat_region   complement(28667. .28773)
                     /rpt_family="PB1D7"
     repeat_region   complement(29357. .29851)
                     /rpt_family="Lx9"
     repeat_region   29853. .30177
                     /rpt_family="ORR1C"
     repeat_region   complement(30179. .30321)
                     /rpt_family="Lx9"
     repeat_region   complement(30371. .30414)
                     /rpt_family="Lx9"
     repeat_region   30503. .30558
                     /rpt_family="B4"
     repeat_region   32026. .32047
                     /rpt_family="(CAGA)n"
     repeat_region   32627. .32806
                     /rpt_family="L1MB3"
     repeat_region   32860. .32990
                     /rpt_family="L1ME2"
     repeat_region   32902. .33121
                     /rpt_family="L1MB3"
     repeat_region   complement(33204. .33502)
                     /rpt_family="MTD"
     repeat_region   complement(35717. .35931)
                     /rpt_family="B4"
     repeat_region   complement(36335. .36679)

  Query Match             90.5%;  Score 19;  DB 10;  Length 276523;
  Best Local Similarity   100.0%;  Pred. No. 4.2e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CATCCTCACCCTTGTCCTC 20
              ||||||||||||||||||||
Db      46248 CATCCTCACCCTTGTCCTC 46230


RESULT 11
BD077670
LOCUS       BD077670                321 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  5'EST of secreted protein expressed in muscles and other mesodermal
            tissues.
ACCESSION   BD077670
VERSION     BD077670.1  GI:22623273
KEYWORDS    JP 2001512016-A/256.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 321)
  AUTHORS   Edwards,J.B.D.M., Duclert,A. and Lacroix,B.
  TITLE     5'EST of secreted protein expressed in muscles and other mesodermal
            tissues
  JOURNAL   Patent: JP 2001512016-A 256 21-AUG-2001;
            GENSET
COMMENT     OS   Homo sapiens (human)
            PN   JP 2001512016-A/256
            PD   21-AUG-2001
            PF   31-JUL-1998 JP 2000505295
            PR   01-AUG-1997 US    08/905134
            PI   JEAN BAPTISTE DUMAS MILNE EDWARDS,AYMERIC DUCLERT,BRUNO  PI
            LACROIX
            PC   C12N15/09,C12N15/09,C07K14/47,C12M1/00,C12N15/00,C12N15/00 CC
             blastn
            CC   identity 98
            CC   region 1. .304
            CC   id HSC26A021
            CC   est
            CC   blastn
            CC   identity 96
            CC   region 1. .158
            CC   id W07871
            CC   est
            CC   blastn
            CC   identity 97
            CC   region 192. .306
            CC   id W07871
            CC   est
            CC   blastn
            CC   identity 96
            CC   region 159. .188
            CC   id W07871
            CC   est
            CC   blastn
            CC   identity 98
            CC   region 144. .280
            CC   id T75539
            CC   est
            CC   blastn
            CC   identity 94
            CC   region 41. .149
            CC   id T75539
            CC   est
            CC   blastn
            CC   identity 100
            CC   region 161. .305
            CC   id H94774
            CC   est
            CC   blastn
            CC   identity 94
            CC   region 10. .151
            CC   id H94774
            CC   est
            CC   blastn
            CC   identity 93
            CC   region 203. .294
            CC   id W89738
            CC   est
            CC   blastn
            CC   identity 91
            CC   region 22. .81
            CC   id W89738
            CC   est
            CC   Von Heijne matrix
            CC   score 6.6
            CC   seq LILVLQLLLRIRR/NR
```

```
            FH   Key             Location/Qualifiers
            FT   misc_feature    16. .319
            FT   misc_feature    17. .174
            FT   misc_feature    205. .319
            FT   misc_feature    174. .203
            FT   misc_feature    169. .305
            FT   misc_feature    64. .172
            FT   misc_feature    175. .319
            FT   misc_feature    24. .165
            FT   misc_feature    228. .319
            FT   misc_feature    43. .102
            FT   sig_peptide     82. .150.
FEATURES             Location/Qualifiers
     source          1. .321
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT       72 a     96 c     84 g     68 t      1 others
ORIGIN

  Query Match            87.6%;  Score 18.4;  DB 6;  Length 321;
  Best Local Similarity  95.0%;  Pred. No. 2e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              ||||||||| ||||||||||
Db        107 TCATCCTCATCCTTGTCCTC 126


RESULT 12
AU028771/c
LOCUS       AU028771                609 bp    DNA     linear   STS 02-MAR-1999
DEFINITION  Rattus norvegicus, OTSUKA clone, OT8.04/02242, microsatellite
            sequence, sequence tagged site.
ACCESSION   AU028771
VERSION     AU028771.1  GI:4518694
KEYWORDS    STS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (sites)
  AUTHORS   Watanabe,T.K., Hishigaki,H., Kanemoto,N., Miyakita-Mizoguchi,A.,
            Oga,K., Okuno,S., Ono,T., Tsuji,A., Hayashi,H., Adachi,M.,
            Yamasaki,Y., Iriye,Y., Takahashi,E., Takagi,T., Nakamura,Y. and
            Tanigami,A.
  TITLE     The large-scale mapping of rat microsatellite markers
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 609)
  AUTHORS   Watanabe,T.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) Takeshi K Watanabe, Otsuka Pharmaceutical
            Co., Ltd., Otsuka GEN Research Institute; 463-10, Kagasuno,
            Kawauchi-cho, Tokushima, Tokushima 771-0192, Japan
            (E-mail:watanabe@otsuka.genome.ad.jp, Tel:81-886-65-2888,
            Fax:81-886-37-1035)
FEATURES             Location/Qualifiers
     source          1. .609
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="OT8.04/02242"
                     /note="OT8.04/02242F=5'-GGGACCCAGAATCAAACCTA-3',
                     OT8.04/02242R=5'-AATTCCACCATACAAGCTAAAGTCA-3'"
BASE COUNT      166 a     99 c    172 g    161 t     11 others
ORIGIN

  Query Match            87.6%;  Score 18.4;  DB 11;  Length 609;
  Best Local Similarity  95.0%;  Pred. No. 1.9e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 CATCCTCACCCTTGTCCTCA 21
              || ||||||||||||||||
Db        317 CACCCTCACCCTTGTCCTCA 298


RESULT 13
AX183446
LOCUS       AX183446                700 bp    DNA     linear   PAT 06-AUG-2001
DEFINITION  Sequence 1199 from Patent WO0142511.
ACCESSION   AX183446
VERSION     AX183446.1  GI:15134759
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Daly,M., Hudson,T.J., Lander,E.S., Rioux,J. and Siminovitch,K.
  TITLE     Ibd-related polymorphisms
  JOURNAL   Patent: WO 0142511-A 1199 14-JUN-2001;
            WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH (US) ; Ellipsis
            Biotherapeutics Corporation (CA)
FEATURES             Location/Qualifiers
     source          1. .700
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      145 a    174 c    226 g    151 t      4 others
ORIGIN

  Query Match            87.6%;  Score 18.4;  DB 6;  Length 700;
  Best Local Similarity  95.0%;  Pred. No. 1.8e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 CATCCTCACCCTTGTCCTCA 21
              || ||||||||||||||||
Db        420 CAGCCTCACCCTTGTCCTCA 439


RESULT 14
VFABAMRCC
LOCUS       VFABAMRCC               993 bp    DNA     linear   PLN 27-APR-1993
DEFINITION  Broad bean (V.faba) BamHI repetitive element, 990 bp family.
ACCESSION   M16855
VERSION     M16855.1  GI:170603
KEYWORDS    BamHI repetitive sequence.
SOURCE      Vicia faba (fava bean)
  ORGANISM  Vicia faba
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae;
            Vicia.
REFERENCE   1  (bases 1 to 993)
  AUTHORS   Kato,A., Iida,Y., Yakura,K. and Tanifuji,S.
  TITLE     Sequence analysis of Vicia faba highly repeated DNA: the BamHI
            repeated sequence families
  JOURNAL   Plant Mol. Biol. 5, 41-53 (1985)
COMMENT     Original source text: Broad bean (V.faba) seed (embryonic axes)
            DNA.
FEATURES             Location/Qualifiers
     source          1. .993
                     /organism="Vicia faba"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3906"
     repeat_region   1. .993
                     /note="BamHI repeat"
BASE COUNT      270 a    260 c    185 g    278 t
ORIGIN      BamHI site.

  Query Match            87.6%;  Score 18.4;  DB 8;  Length 993;
  Best Local Similarity  95.0%;  Pred. No. 1.7e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
```

```
                  |||||||||| ||||||||
Db        911 TCATCCTCACCTTTGTCCTC 930


RESULT 15
AB020881
LOCUS       AB020881               1227 bp    mRNA    linear   PRI 17-FEB-1999
DEFINITION  Homo sapiens mRNA for TBP-like protein, complete cds.
ACCESSION   AB020881
VERSION     AB020881.1  GI:4579752
KEYWORDS    TBP-like protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ohbayashi,T., Kishimoto,T., Makino,Y., Shimada,M., Nakadai,T.,
            Aoki,T., Kawata,T., Niwa,S. and Tamura,T.
  TITLE     Isolation of cDNA, chromosome mapping, and expression of the human
            TBP-like protein
  JOURNAL   Biochem. Biophys. Res. Commun. 255 (1), 137-142 (1999)
  MEDLINE   99185310
   PUBMED   10082669
REFERENCE   2  (bases 1 to 1227)
  AUTHORS   Ohbayashi,T., Kishimoto,T., Makino,Y. and Tamura,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1998) Toshihiko Kishimoto, Sumitomo Electric
            Industries, Biomedical R&D Dept.; 1, Taya-cho, Sakae-ku, Yokohama,
            Kanagawa 244-8588, Japan (E-mail:tkishmt@opele.sei.co.jp,
            Tel:81-45-853-7275, Fax:81-45-853-3528)
FEATURES             Location/Qualifiers
     source          1. .1227
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6q22.1-22.3"
     CDS             223. .783
                     /note="TLP"
                     /codon_start=1
                     /product="TBP-like protein"
                     /protein_id="BAA75218.1"
                     /db_xref="GI:4579753"
                     /translation="MDADSDVALDILITNVVCVFRTRCHLNLRKIALEGANVIYKRDV
                     GKVLMKLRKPRITATIWSSGKIICTGATSEEEAKFGARRLARSLQKLGFQVIFTDFKV
                     VNVLAVCNMPFEIRLPEFTKNNRPHASYEPELHPAVCYRIKSLRATLQIFSTGSITVT
                     GPNVKAVATAVEQIYPFVFESRKEIL"
     polyA_site      1227
                     /note="19 A nucleotides"
BASE COUNT      357 a    247 c    254 g    369 t
ORIGIN

  Query Match             87.6%;  Score 18.4;  DB 9;  Length 1227;
  Best Local Similarity   95.0%;  Pred. No. 1.7e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              |||||||||| ||||||||||
Db         82 TCATCCTCATCCTTGTCCTC 101


Search completed: February 19, 2004, 22:22:00
Job time : 1601.5 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 19:59:22 ; Search time 175.5 Seconds
(without alignments)
323.010 Million cell updates/sec

Title: US-10-085-108-21_COPY_175_195
Perfect score: 21
Sequence: 1 TCATCCTCACCCTTGTCCTCA 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21	100.0	21	25	ABX95021	Human MAGE-C3 expr
c	2	21	100.0	424	22	ABA51188	Human breast cell
c	3	21	100.0	424	22	ABA69192	Human foetal liver
c	4	21	100.0	424	22	ABA36116	Probe #14582 for g
c	5	21	100.0	424	22	AAK17490	Human brain expres
c	6	21	100.0	424	22	AAK43299	Human bone marrow
c	7	21	100.0	424	22	AAI24071	Probe #14004 for g
c	8	21	100.0	424	22	AAI49367	Probe #18053 used
c	9	21	100.0	424	22	AAI09651	Probe #9642 used t
c	10	21	100.0	424	23	ABS42926	Human liver single
c	11	21	100.0	424	24	ABS17378	Human genome-deriv
	12	21	100.0	1041	25	ABX95006	cDNA encoding huma
	13	21	100.0	7806	23	AAS88354	DNA encoding novel
	14	18.4	87.6	321	20	AAX41050	Human secreted pro
	15	18.4	87.6	700	22	AAH93187	Human inflammatory
c	16	18.4	87.6	772	22	AAH34669	Human colon cancer
	17	18.4	87.6	852	22	AAS26209	Human cDNA encodin
	18	18.4	87.6	852	25	ABX73550	Human novel polynu
	19	18.4	87.6	1125	21	AAZ57867	Protein regulating
	20	18.4	87.6	1246	21	AAA97486	Human transcriptio
c	21	18.4	87.6	1697	24	ABL55652	Signal peptidase 2
c	22	18.4	87.6	2043	21	AAA75119	cDNA encoding a hu
c	23	18.4	87.6	2521	23	AAS67756	DNA encoding novel
c	24	17.8	84.8	1251	21	AAA09347	Human MBP-calretic
c	25	17.8	84.8	1890	10	AAN91471	1.9 kb Ro (Ro/SSA)
c	26	17.8	84.8	1899	24	ABV73173	Human calreticulin
c	27	17.8	84.8	1899	24	ABK11662	DNA encoding human
c	28	17.8	84.8	1899	25	ABT17040	Human MP21 gene CR
c	29	17.8	84.8	1920	24	AAD29931	Human calreticulin
c	30	17.8	84.8	1924	24	AAS95010	Human DNA sequence
c	31	17.8	84.8	1958	21	AAA09346	Human MBP-calretic
c	32	17.8	84.8	1958	24	AAD39469	Human calreticulin
c	33	17.8	84.8	2086	23	AAS86143	DNA encoding novel
c	34	17.8	84.8	4054	24	ABN60011	Novel human coding
c	35	17.8	84.8	110096	24	ABN95044	Gene #1542 used to
c	36	17.4	82.9	432	22	ABA45379	Human breast cell
c	37	17.4	82.9	432	22	ABA55869	Human foetal liver
c	38	17.4	82.9	432	22	ABA25543	Probe #4009 for ge
c	39	17.4	82.9	432	22	AAK04087	Human brain expres
c	40	17.4	82.9	432	22	AAK29569	Human bone marrow
c	41	17.4	82.9	432	22	AAI14143	Probe #4076 for ge
c	42	17.4	82.9	432	22	AAI35528	Probe #4214 used t
c	43	17.4	82.9	432	22	AAI03991	Probe #3982 used t
c	44	17.4	82.9	432	23	ABS29199	Human liver single
c	45	17.4	82.9	432	24	ABS04124	Human genome-deriv

ALIGNMENTS

RESULT 1
ABX95021
ID ABX95021 standard; DNA; 21 BP.
XX
AC ABX95021;
XX
DT 05-JUN-2003 (first entry)
XX
DE Human MAGE-C3 expression pattern anlaysis RT-PCR sense primer.
XX
KW TRAP; ss; tumour rejection antigen precursor; cytolytic T-cell; CTL;
KW tumour; seminoma; bladder transitional-cell carcinoma; NSCLC; adaptor;
KW head-and-neck squamous-cell carcinoma; breast carcinoma; sarcoma;
KW cutaneous melanoma; nonsmall cell lung cancer; RT-PCR; primer; MAGE-C3;
KW human; reverse transcription.
XX
OS Homo sapiens.
XX
PN US2002176865-A1.
XX
PD 28-NOV-2002.
XX
PF 01-MAR-2002; 2002US-0085108.
XX
PR 09-FEB-2000; 2000US-0501104.
PR 25-APR-1997; 97US-0845528.
PR 24-APR-1998; 98US-0066281.
PR 17-DEC-1999; 99US-0468433.
XX
PA (LUCA/) LUCAS S.
PA (BOON/) BOON-FALLEUR T.

XX
PI Lucas S, Boon-Falleur T;
XX
DR WPI; 2003-328468/31.
XX
PT Novel isolated nucleic acid encoding tumor rejection antigen precursor
PT MAGE-C3, MAGE-B5, or MAGE-B6, useful as diagnostic probes to determine
PT presence of abnormal e.g., tumor cells expressing MAGE-C1, MAGE-B5 or
PT MAGE-B6 -
XX
PS Example 12; Page 13; 59pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule which encodes
CC a tumour rejection antigen precursor (TRAP) having an amino acid sequence
CC of a TRAP encoded by a fully defined MAGE-C3, MAGE-B5, or MAGE-B6
CC polynucleotide sequence. Also disclosed is a method which is useful for
CC determining presence of cytolytic T-cells specific for complexes of human
CC leukocyte antigen (HLA) and a peptide derived from the nucleic acid in a
CC cytotoxic T-lymphocyte (CTL)-containing sample. The nucleic acid is
CC useful as a diagnostic probe to determine the presence of abnormal
CC (tumour) cells such as seminoma, bladder transitional-cell carcinoma,
CC head-and-neck squamous-cell carcinoma, breast carcinoma, sarcoma,
CC cutaneous melanoma or nonsmall cell lung cancer (NSCLC) which express
CC MAGE-C1, MAGE-B5 or MAGE-B6. The nucleic acid is useful for diagnosing a
CC disorder characterised by expression of MAGE-C1, MAGE-B5 or MAGE-B6 TRAPs
CC or tumour rejection antigens (TRAs). The present sequence represents the
CC human MAGE-C3 expression pattern anlaysis reverse transcription (RT)-PCR
CC sense primer.
XX
SQ Sequence 21 BP; 3 A; 10 C; 1 G; 7 T; 0 other;

Query Match 100.0%; Score 21; DB 25; Length 21;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGTCCTCA 21
||||||||||||||||||||||
Db 1 TCATCCTCACCCTTGTCCTCA 21

RESULT 2
ABA51188/c
ID ABA51188 standard; DNA; 424 BP.
XX
AC ABA51188;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #9883.
XX
KW Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
PS Claim 4; SEQ ID NO 9883; 327pp + sequence listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 424 BP; 108 A; 94 C; 129 G; 93 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 424;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGTCCTCA 21
||||||||||||||||||||||
Db 411 TCATCCTCACCCTTGTCCTCA 391

RESULT 3
ABA69192/c
ID ABA69192 standard; DNA; 424 BP.
XX
AC ABA69192;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #17497.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human fetal liver -
XX
PS Claim 4; SEQ ID NO 17497; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 424 BP; 108 A; 94 C; 129 G; 93 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 424;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGTCCTCA 21
||||||||||||||||||||
Db 411 TCATCCTCACCCTTGTCCTCA 391

RESULT 4
ABA36116/c
ID ABA36116 standard; DNA; 424 BP.
XX
AC ABA36116;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #14582 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 4; SEQ ID No 14582; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 424 BP; 108 A; 94 C; 129 G; 93 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 424;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGTCCTCA 21
||||||||||||||||||||
Db 411 TCATCCTCACCCTTGTCCTCA 391

RESULT 5
AAK17490/c
ID AAK17490 standard; DNA; 424 BP.
XX
AC AAK17490;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 17481.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 17481; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 424 BP; 108 A; 94 C; 129 G; 93 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 424;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGTCCTCA 21

```
                    ||||||||||||||||||||
Db        411 TCATCCTCACCCTTGTCCTCA 391
```

RESULT 6
AAK43299/c
ID AAK43299 standard; DNA; 424 BP.
XX
AC AAK43299;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 17856.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 17856; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 424 BP; 108 A; 94 C; 129 G; 93 T; 0 other;

  Query Match   100.0%; Score 21; DB 22; Length 424;
  Best Local Similarity   100.0%; Pred. No. 16;
  Matches   21; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          1 TCATCCTCACCCTTGTCCTCA 21
              ||||||||||||||||||||
Db        411 TCATCCTCACCCTTGTCCTCA 391
```

RESULT 7
AAI24071/c
ID AAI24071 standard; DNA; 424 BP.
XX
AC AAI24071;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #14004 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 14004; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 424 BP; 108 A; 94 C; 129 G; 93 T; 0 other;

  Query Match   100.0%; Score 21; DB 22; Length 424;
  Best Local Similarity   100.0%; Pred. No. 16;
  Matches   21; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          1 TCATCCTCACCCTTGTCCTCA 21
              ||||||||||||||||||||
Db        411 TCATCCTCACCCTTGTCCTCA 391
```

RESULT 8
AAI49367/c
ID AAI49367 standard; DNA; 424 BP.
XX
AC AAI49367;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #18053 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 18053; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 424 BP; 108 A; 94 C; 129 G; 93 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 424;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGTCCTCA 21
||||||||||||||||||||
Db 411 TCATCCTCACCCTTGTCCTCA 391

RESULT 9
AAI09651/c
ID AAI09651 standard; DNA; 424 BP.
XX
AC AAI09651;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #9642 used to measure gene expression in human breast sample.
XX
KW Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 25; SEQ ID No 9642; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 424 BP; 108 A; 94 C; 129 G; 93 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 424;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGTCCTCA 21
||||||||||||||||||||
Db 411 TCATCCTCACCCTTGTCCTCA 391

RESULT 10
ABS42926/c
ID ABS42926 standard; DNA; 424 BP.
XX
AC ABS42926;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver single exon probe, SEQ ID No 17916.
XX
KW Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00664.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488898/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analysing gene expression in human adult liver -
XX
PS Claim 4; SEQ ID No 17916; 658pp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult

CC liver. (I) may be used for predicting, measuring and displaying gene
CC expression in samples derived from human adult liver. The genes
CC identified may be involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
CC is associated with coronary heart disease. ABS25011-ABS51005 represent
CC human liver single exon nucleic acid probes of the invention.
CC Note: The sequence information for this patent does not appear in the
CC printed specification but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 424 BP; 108 A; 94 C; 129 G; 93 T; 0 other;

Query Match 100.0%; Score 21; DB 23; Length 424;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGTCCTCA 21
||||||||||||||||||||
Db 411 TCATCCTCACCCTTGTCCTCA 391

RESULT 11
ABS17378/c
ID ABS17378 standard; DNA; 424 BP.
XX
AC ABS17378;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe ORF from lung SEQ ID No 17369.
XX
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesis; pulmonary hypertension;
KW hyaline membrane disease; open reading frame; ORF.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX
PS Claim 4; SEQ ID No 17369; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomtosis,
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesis, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a single exon
CC probe open reading frame of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 424 BP; 108 A; 94 C; 129 G; 93 T; 0 other;

Query Match 100.0%; Score 21; DB 24; Length 424;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGTCCTCA 21
||||||||||||||||||||
Db 411 TCATCCTCACCCTTGTCCTCA 391

RESULT 12
ABX95006
ID ABX95006 standard; cDNA; 1041 BP.
XX
AC ABX95006;
XX
DT 05-JUN-2003 (first entry)
XX
DE cDNA encoding human tumour rejection antigen precursor, MAGE-C3.
XX
KW TRAP; ss; tumour rejection antigen precursor; cytolytic T-cell; CTL;
KW tumour; seminoma; bladder transitional-cell carcinoma; NSCLC; adaptor;
KW head-and-neck squamous-cell carcinoma; breast carcinoma; sarcoma;
KW cutaneous melanoma; nonsmall cell lung cancer; gene; MAGE-C3; human;
KW chromosome Xq27.1-Xq27.3.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1041
FT /*tag= a
FT /product= "MAGE-C3"
XX
PN US2002176865-A1.
XX
PD 28-NOV-2002.

```
XX
PF   01-MAR-2002; 2002US-0085108.
XX
PR   09-FEB-2000; 2000US-0501104.
PR   25-APR-1997;   97US-0845528.
PR   24-APR-1998;   98US-0066281.
PR   17-DEC-1999;   99US-0468433.
XX
PA   (LUCA/) LUCAS S.
PA   (BOON/) BOON-FALLEUR T.
XX
PI   Lucas S,  Boon-Falleur T;
XX
DR   WPI; 2003-328468/31.
DR   P-PSDB; ABU08932.
XX
PT   Novel isolated nucleic acid encoding tumor rejection antigen precursor
PT   MAGE-C3, MAGE-B5, or MAGE-B6, useful as diagnostic probes to determine
PT   presence of abnormal e.g., tumor cells expressing MAGE-C1, MAGE-B5 or
PT   MAGE-B6  -
XX
PS   Claim 1; Fig 5; 59pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a tumour rejection antigen precursor (TRAP) having an amino acid sequence
CC   of a TRAP encoded by a fully defined MAGE-C3, MAGE-B5, or MAGE-B6
CC   polynucleotide sequence. Also disclosed is a method which is useful for
CC   determining presence of cytolytic T-cells specific for complexes of human
CC   leukocyte antigen (HLA) and a peptide derived from the nucleic acid in a
CC   cytotoxic T-lymphocyte (CTL)-containing sample. The nucleic acid is
CC   useful as a diagnostic probe to determine the presence of abnormal
CC   (tumour) cells such as seminoma, bladder transitional-cell carcinoma,
CC   head-and-neck squamous-cell carcinoma, breast carcinoma, sarcoma,
CC   cutaneous melanoma or nonsmall cell lung cancer (NSCLC) which express
CC   MAGE-C1, MAGE-B5 or MAGE-B6. The nucleic acid is useful for diagnosing a
CC   disorder characterised by expression of MAGE-C1, MAGE-B5 or MAGE-B6 TRAPs
CC   or tumour rejection antigens (TRAs). The present sequence represents the
CC   cDNA of the gene encoding the human tumour rejection antigen precursor,
CC   MAGE-C3, which is located on chromosome Xq27.1-Xq27.3.
XX
SQ   Sequence 1041 BP; 242 A; 283 C; 242 G; 274 T; 0 other;

  Query Match             100.0%;  Score 21;  DB 25;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              |||||||||||||||||||||
Db        175 TCATCCTCACCCTTGTCCTCA 195


RESULT 13
AAS88354
ID   AAS88354 standard; cDNA; 7806 BP.
XX
AC   AAS88354;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #24158.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG24167.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 24158; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 7806 BP; 1755 A; 2185 C; 1760 G; 2106 T; 0 other;

  Query Match             100.0%;  Score 21;  DB 23;  Length 7806;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              |||||||||||||||||||||
Db       3435 TCATCCTCACCCTTGTCCTCA 3455


RESULT 14
AAX41050
ID   AAX41050 standard; cDNA; 321 BP.
XX
AC   AAX41050;
XX
DT   18-JUN-1999  (first entry)
XX
DE   Human secreted protein 5' EST SEQ ID NO: 262.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS   Homo sapiens.
XX
PN   WO9906554-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01238.
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 21:05:58 ; Search time 1571 Seconds
(without alignments)
324.885 Million cell updates/sec

Title: US-10-085-108-21_COPY_175_195
Perfect score: 21
Sequence: 1 TCATCCTCACCCTTGTCCTCA 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match %	Length	DB	ID	Description
	1	21	100.0	1062	12	BM547686	BM547686 AGENCOURT
	2	19.4	92.4	511	13	BQ496083	BQ496083 EST05312
c	3	19.4	92.4	592	29	BZ298653	BZ298653 CG4508.r1
c	4	19.4	92.4	600	28	AZ246440	AZ246440 RPCI-23-9
c	5	19.4	92.4	736	28	AZ247894	AZ247894 RPCI-23-9
	6	18.4	87.6	186	14	CB118019	CB118019 K-EST0164
c	7	18.4	87.6	237	9	AV015025	AV015025 AV015025
	8	18.4	87.6	273	9	AA298913	AA298913 EST114541
	9	18.4	87.6	280	10	BB032507	BB032507 BB032507
	10	18.4	87.6	291	9	AI220795	AI220795 qg02e10.x
	11	18.4	87.6	294	10	BB488898	BB488898 BB488898
	12	18.4	87.6	308	14	F07446	F07446 HSC26A021 n
	13	18.4	87.6	324	9	AI025315	AI025315 ov74e08.x
	14	18.4	87.6	339	12	BM843625	BM843625 K-EST0121
	15	18.4	87.6	379	14	CB119380	CB119380 K-EST0166
	16	18.4	87.6	382	14	CB118879	CB118879 K-EST0165
	17	18.4	87.6	386	9	AI285916	AI285916 qh97b07.x
	18	18.4	87.6	412	14	T75539	T75539 yd63e08.r1
	19	18.4	87.6	416	28	AZ043988	AZ043988 RPCI-23-3
	20	18.4	87.6	422	9	AA297902	AA297902 EST113464
	21	18.4	87.6	422	10	BE246429	BE246429 TCBAP1E46
	22	18.4	87.6	425	10	BF919684	BF919684 RC6-NT015
	23	18.4	87.6	434	13	BU947841	BU947841 io54c12.y
	24	18.4	87.6	441	14	H94774	H94774 yw58e07.r1
	25	18.4	87.6	463	13	BU657956	BU657956 cl30d02.z
	26	18.4	87.6	463	14	W07871	W07871 zb04h06.r1
	27	18.4	87.6	465	14	CB119687	CB119687 K-EST0166
	28	18.4	87.6	466	9	AA281228	AA281228 zs94c11.r
	29	18.4	87.6	476	12	BM842487	BM842487 K-EST0119
	30	18.4	87.6	480	14	CB120628	CB120628 K-EST0167
	31	18.4	87.6	486	9	AW161234	AW161234 au70b11.y
	32	18.4	87.6	489	10	BF059384	BF059384 7k57h01.x
	33	18.4	87.6	497	13	BU656535	BU656535 cl12f04.z
	34	18.4	87.6	516	13	BU659222	BU659222 cl45b04.z
	35	18.4	87.6	519	13	BX280844	BX280844 BX280844
	36	18.4	87.6	530	28	AZ988048	AZ988048 2M0270A22
	37	18.4	87.6	539	9	AW629726	AW629726 hh70a09.y
	38	18.4	87.6	546	10	BG494919	BG494919 602540937
	39	18.4	87.6	551	9	AI791890	AI791890 nl95c11.y
	40	18.4	87.6	552	10	BG494575	BG494575 602540623
	41	18.4	87.6	553	13	BU661680	BU661680 cl75d02.z
	42	18.4	87.6	554	10	BF669343	BF669343 602120628
	43	18.4	87.6	555	29	BX241672	BX241672 Danio rer
	44	18.4	87.6	579	10	BG528372	BG528372 602557790
	45	18.4	87.6	585	9	AW247260	AW247260 2820491.5

ALIGNMENTS

RESULT 1
BM547686
LOCUS BM547686 1062 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT_6507236 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5727806 5', mRNA sequence.
ACCESSION BM547686
VERSION BM547686.1 GI:18781667
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1062)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12722 row: h column: 15
High quality sequence stop: 631.

```
FEATURES             Location/Qualifiers
     source          1. .1062
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5727806"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_124"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: EcoRV
                     (destroyed); Site_2: NotI; RNA source male hippocampus,
                     age 27. Library is oligo-dT primed and directionally
                     cloned (EcoRV site is destroyed upon cloning). Average
                     insert size 1.4 kb, insert size range 0.9-4 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 012."
BASE COUNT      246 a    330 c    254 g    232 t
ORIGIN

  Query Match             100.0%;  Score 21;  DB 12;  Length 1062;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              |||||||||||||||||||||
Db        700 TCATCCTCACCCTTGTCCTCA 720


RESULT 2
BQ496083
LOCUS       BQ496083                 511 bp    mRNA    linear   EST 31-OCT-2002
DEFINITION  EST05312 Pb0001 Paracoccidioides brasiliensis cDNA, mRNA sequence.
ACCESSION   BQ496083
VERSION     BQ496083.1  GI:24446451
KEYWORDS    EST.
SOURCE      Paracoccidioides brasiliensis
  ORGANISM  Paracoccidioides brasiliensis
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; mitosporic Onygenales; Paracoccidioides.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Goldman,G.H., Marques,E.R., Ribeiro,D.C.D., Bernardes,L.A., Puccia
            ,R., Travassos,L.R., Nobrega,F.G., Nobrega,M.P., Savoldi-Barbosa,M.
            , Semighini,C.P. and Goldman,M.H.
  TITLE     The Paracoccidioides brasiliensis EST genome project
  JOURNAL   Eukaryotic Cell,  (2002) In press
COMMENT     Contact: Gustavo Henrique Goldman
            Laboratory of Molecular Biology
            Universidade de Sao Paulo - USP - FCFRP
            Av do Cafe S/N, CEP: 14040-903, Ribeirao Preto - SP, Brazil
            Email: ggoldman@usp.br.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Paracoccidioides brasiliensis"
                     /mol_type="mRNA"
                     /db_xref="taxon:121759"
                     /clone_lib="Pb0001"
BASE COUNT      115 a    177 c     65 g    154 t
ORIGIN

  Query Match             92.4%;  Score 19.4;  DB 13;  Length 511;
  Best Local Similarity   95.2%;  Pred. No. 1.9e+03;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              |||||||||||||||||| ||
Db        290 TCATCCTCACCCTTGTCCCCA 310


RESULT 3
BZ298653/c
LOCUS       BZ298653                 592 bp    DNA     linear   GSS 31-OCT-2002
DEFINITION  CG4508.r1 Candida glabrata Random Genomic Library Candida glabrata
            genomic clone CG4508, genomic survey sequence.
ACCESSION   BZ298653
VERSION     BZ298653.1  GI:24442019
KEYWORDS    GSS.
SOURCE      Candida glabrata
  ORGANISM  Candida glabrata
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
  TITLE     Evidence from comparative genomics for a complete sexual cycle in
            the 'asexual' pathogenic yeast Candida glabrata
  JOURNAL   Genome Biol. 4 (2), R10 (2003)
  MEDLINE   22508158
   PUBMED   12620120
COMMENT     Contact: Wong S
            Department of Genetics, Smurfit Institute
            Trinity College Dublin
            Dublin 2, Ireland
            Tel: 353 1 6082319
            Fax: 353 1 6798558
            Email: swong@tcd.ie
            Class: plasmid ends.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Candida glabrata"
                     /mol_type="genomic DNA"
                     /strain="CBS 138"
                     /db_xref="taxon:5478"
                     /clone="CG4508"
                     /clone_lib="Candida glabrata Random Genomic Library"
BASE COUNT      209 a     95 c    160 g    128 t
ORIGIN

  Query Match             92.4%;  Score 19.4;  DB 29;  Length 592;
  Best Local Similarity   95.2%;  Pred. No. 2e+03;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              |||||||||||| |||||||||
Db        374 TCATCCTCACCTTTGTCCTCA 354


RESULT 4
AZ246440/c
LOCUS       AZ246440                 600 bp    DNA     linear   GSS 15-JUN-2000
DEFINITION  RPCI-23-91H10.TV RPCI-23 Mus musculus genomic clone RPCI-23-91H10,
            genomic survey sequence.
ACCESSION   AZ246440
VERSION     AZ246440.1  GI:8559637
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI-23-91H10.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
```

```
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 91  row: H  column: 10
            Seq primer: T7
            Class: BAC ends.
FEATURES                Location/Qualifiers
     source             1. .600
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="RPCI-23-91H10"
                        /sex="Female"
                        /lab_host="DH10B"
                        /clone_lib="RPCI-23"
                        /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                        EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                        brain genomic DNA was isolated and  partially digested
                        with a combination of EcoRI and EcoRI Methylase.   Size
                        selected DNA was cloned into the pBACe3.6 vector at the
                        EcoRI  sites.  The ligation products were transformed into
                        DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT      161 a    134 c    166 g    139 t
ORIGIN

  Query Match             92.4%;  Score 19.4;  DB 28;  Length 600;
  Best Local Similarity   95.2%;  Pred. No. 2e+03;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              || |||||||||||||||||||
Db        370 TCCTCCTCACCCTTGTCCTCA 350


RESULT 5
AZ247894/c
LOCUS       AZ247894                 736 bp    DNA     linear   GSS 15-JUN-2000
DEFINITION  RPCI-23-91L10.TV RPCI-23 Mus musculus genomic clone RPCI-23-91L10,
            genomic survey sequence.
ACCESSION   AZ247894
VERSION     AZ247894.1  GI:8561097
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 736)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI-23-91L10.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 91  row: L  column: 10
            Seq primer: T7
            Class: BAC ends.
FEATURES                Location/Qualifiers
     source             1. .736
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="RPCI-23-91L10"
                        /sex="Female"
                        /lab_host="DH10B"
                        /clone_lib="RPCI-23"
                        /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                        EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                        brain genomic DNA was isolated and  partially digested
                        with a combination of EcoRI and EcoRI Methylase.   Size
                        selected DNA was cloned into the pBACe3.6 vector at the
                        EcoRI  sites.  The ligation products were transformed into
                        DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT      216 a    158 c    197 g    165 t
ORIGIN

  Query Match             92.4%;  Score 19.4;  DB 28;  Length 736;
  Best Local Similarity   95.2%;  Pred. No. 2.1e+03;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              || |||||||||||||||||||
Db        373 TCCTCCTCACCCTTGTCCTCA 353


RESULT 6
CB118019
LOCUS       CB118019                 186 bp    mRNA    linear   EST 28-JAN-2003
DEFINITION  K-EST0164201 B1T694954 Homo sapiens cDNA clone B1T694954-9-A02 5',
            mRNA sequence.
ACCESSION   CB118019
VERSION     CB118019.1  GI:27943826
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens .
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 9  row: A  column: 02
            High quality sequence stop: 186.
FEATURES                Location/Qualifiers
     source             1. .186
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="B1T694954-9-A02"
                        /sex="M"
                        /lab_host="Top10F'"
                        /clone_lib="B1T694954"
                        /note="Organ: Brain; Vector: pCNS-D2; Site_1: EcoRI;
                        Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
                        bacterial alkaline phosphatase (BAP) and then decapped
                        with tabacco acid pyrophosphatase (TAP). The decapped
                        intact mRNA was ligated with DNA-RNA linker including
                        EcoRI site by treatment of T4 RNA ligase and the first
                        strand cDNA was synthesized from oligo dT-selected mRNA by
                        priming with dT-tailed vector. The dT-tailed vector was
                        adjusted to have about 60nt. The cDNA vector was
                        circularized with E. coli DNA ligase after digestion of
                        EcoRI which site is also included in vector. An RNA strand
                        converted to a DNA strand by Okayama-Berg method. The
```

obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

BASE COUNT 34 a 60 c 57 g 35 t
ORIGIN

Query Match 87.6%; Score 18.4; DB 14; Length 186;
Best Local Similarity 95.0%; Pred. No. 3.6e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGTCCTC 20
||||||||| ||||||||||
Db 165 TCATCCTCATCCTTGTCCTC 184

RESULT 7
AV015025/c
LOCUS AV015025 237 bp mRNA linear EST 28-AUG-1999
DEFINITION AV015025 Mus musculus 18-day embryo C57BL/6J Mus musculus cDNA clone 1110057J14, mRNA sequence.
ACCESSION AV015025
VERSION AV015025.1 GI:4792017
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 237)
AUTHORS Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K., Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs
JOURNAL Unpublished
COMMENT Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .237
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1110057J14"
/sex="mixed"
/dev_stage="18-day embryo"
/clone_lib="Mus musculus 18-day embryo C57BL/6J"
BASE COUNT 57 a 50 c 63 g 67 t
ORIGIN

Query Match 87.6%; Score 18.4; DB 9; Length 237;
Best Local Similarity 95.0%; Pred. No. 3.8e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGTCCTC 20
|| ||||||||||||||||||
Db 106 TCCTCCTCACCCTTGTCCTC 87

RESULT 8
AA298913
LOCUS AA298913 273 bp mRNA linear EST 18-APR-1997
DEFINITION EST114541 Testis tumor Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA298913
VERSION AA298913.1 GI:1951276
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 273)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
PUBMED 7566098
COMMENT Other_ESTs: THC173596
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .273
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="ATCC (inhost):191897"
/db_xref="taxon:9606"
/sex="male"
/dev_stage="adult"
/clone_lib="Testis tumor"
/note="Organ: testis; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 57 a 92 c 70 g 49 t 5 others
ORIGIN

Query Match 87.6%; Score 18.4; DB 9; Length 273;
Best Local Similarity 95.0%; Pred. No. 3.9e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGTCCTC 20
||||||||| ||||||||||
Db 111 TCATCCTCATCCTTGTCCTC 130

RESULT 9
BB032507

```
LOCUS       BB032507                280 bp    mRNA    linear   EST 23-JUN-2000
DEFINITION  BB032507 RIKEN full-length enriched, adult male thymus Mus musculus
            cDNA clone 5830471I16 3' similar to U78167 Rattus norvegicus
            cAMP-regulated guanine nucleotide exchange factor I (cAMP-GEFI)
            mRNA, mRNA sequence.
ACCESSION   BB032507
VERSION     BB032507.1  GI:8387258
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 280)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .280
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="5830471I16"
                     /sex="male"
                     /tissue_type="thymus"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male thymus"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI."
BASE COUNT       52 a     75 c     68 g     85 t
ORIGIN

  Query Match             87.6%;  Score 18.4;  DB 10;  Length 280;
  Best Local Similarity   95.0%;  Pred. No. 3.9e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              |||||||||||| ||||||
Db         37 TCATCCTCACCCTCGTCCTC 56


RESULT 10
AI220795
LOCUS       AI220795                291 bp    mRNA    linear   EST 29-NOV-1998
DEFINITION  qg02e10.x1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA
            clone IMAGE:1758378 3', mRNA sequence.
ACCESSION   AI220795
VERSION     AI220795.1  GI:3802998
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1072   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 286.
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1758378"
                     /dev_stage="two placentae: one from 8 weeks and another
                     from 9 weeks post conception"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares_placenta_8to9weeks_2NbHP8to9W"
                     /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo."
BASE COUNT       59 a     96 c     82 g     54 t
ORIGIN

  Query Match             87.6%;  Score 18.4;  DB 9;  Length 291;
  Best Local Similarity   95.0%;  Pred. No. 4e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              |||||||| ||||||||||
Db        114 TCATCCTCATCCTTGTCCTC 133
```

```
RESULT 11
BB488898
LOCUS       BB488898                294 bp    mRNA    linear   EST 23-JUL-2000
DEFINITION  BB488898 RIKEN full-length enriched, 13 days embryo stomach Mus
            musculus cDNA clone D530004H21 3', mRNA sequence.
ACCESSION   BB488898
VERSION     BB488898.1  GI:9406507
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 294)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .294
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="D530004H21"
                     /tissue_type="stomach"
                     /dev_stage="13 days embryo"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 13 days embryo
                     stomach"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT       81 a     64 c     41 g    108 t
ORIGIN

  Query Match            87.6%;  Score 18.4;  DB 10;  Length 294;
  Best Local Similarity  95.0%;  Pred. No. 4e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              ||||||||| ||||||||||
Db         24 TCATCCTCATCCTTGTCCTC 43


RESULT 12
F07446
LOCUS       F07446                  308 bp    mRNA    linear   EST 20-FEB-1995
DEFINITION  HSC26A021 normalized infant brain cDNA Homo sapiens cDNA clone
            c-26a02, mRNA sequence.
ACCESSION   F07446
VERSION     F07446.1  GI:673106
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 308)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
            ,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
            Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
   PUBMED   7757816
COMMENT     Contact: Genethon
            Genexpress-Genethon
            Genethon Centre de recherche sur le Genome Humain
            1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
            Tel: 33169472800
            Fax: 33160778698
            Email: genexpress@genethon.fr
            Single read.
            Genexpress_library_idt: C; Genexpress_sequence_idt: y1c-26a02
            Seq primer: (-21)M13_universal.
FEATURES             Location/Qualifiers
     source          1. .308
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="c-26a02"
                     /sex="Female"
                     /tissue_type="total brain"
                     /dev_stage="3 months old"
                     /clone_lib="normalized infant brain cDNA"
                     /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                     Site_2: NotI; sex=Female; dev_stage=3 months old;
                     isolate=muscular atrophy patient;  tissue_type=total brain
                     ;  total mRNA was oligo-(dT) primed and directionally
                     cloned 5' -> 3' into the  HindIII -> NotI sites of the
                     lafmid BA vector. Clone library from B.Soares,  Psychiatry
                     Dept. Columbia University, USA.  Normalization_method:
                     Bento Soares, P.N.A.S in press"
BASE COUNT       68 a     89 c     79 g     69 t      3 others
ORIGIN

  Query Match            87.6%;  Score 18.4;  DB 14;  Length 308;
  Best Local Similarity  95.0%;  Pred. No. 4e+03;
```

```
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              |||||||| ||||||||||
Db         92 TCATCCTCATCCTTGTCCTC 111


RESULT 13
AI025315
LOCUS       AI025315                324 bp    mRNA    linear   EST 27-AUG-1998
DEFINITION  ov74e08.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1643078
            3' similar to contains element MER22 repetitive element ;, mRNA
            sequence.
ACCESSION   AI025315
VERSION     AI025315.1  GI:3240928
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 324)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 383   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .324
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1643078"
                     /sex="male"
                     /lab_host="DH10B"
                     /clone_lib="Soares_testis_NHT"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       51 a     93 c     82 g     98 t
ORIGIN

  Query Match             87.6%;  Score 18.4;  DB 9;  Length 324;
  Best Local Similarity   95.0%;  Pred. No. 4.1e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 CATCCTCACCCTTGTCCTCA 21
              |||||||| ||||||||||
Db        262 CATCCTCATCCTTGTCCTCA 281


RESULT 14
BM843625
LOCUS       BM843625                339 bp    mRNA    linear   EST 06-MAR-2002
DEFINITION  K-EST0121571 S12SNU216 Homo sapiens cDNA clone S12SNU216-71-E10 5',
            mRNA sequence.
ACCESSION   BM843625
VERSION     BM843625.1  GI:19200034
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 71  row: E  column: 10
            High quality sequence stop: 339.
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="S12SNU216-71-E10"
                     /sex="F"
                     /tissue_type="Lymph node"
                     /cell_type="Epithelial"
                     /cell_line="SNU-216"
                     /lab_host="Top10F'"
                     /clone_lib="S12SNU216"
                     /note="Organ: Stomach; Vector: pCNS; Site_1: EcoRI;
                     Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
                     bacterial alkaline phosphatase (BAP) and then decapped
                     with tabacco acid pyrophosphatase (TAP). The decapped
                     intact mRNA was ligated with DNA-RNA linker including EcoR
                     I site by treatment of T4 RNA ligase and the first strand
                     cDNA was synthesized from oligo dT-selected mRNA by
                     priming with dT-tailed vector. The dT-tailed vector was
                     adjusted to have about 60nt. The cDNA vector was
                     circularized with E. coli DNA ligase after digestion of
                     EcoRI which site is also included in vector. An RNA strand
                     converted to a DNA strand by Okayama-Berg method. The
                     obtained cDNA vectors were used for transformation of
                     competent cells E. coli Top10F' by electroporation method.
                     The cDNA libraries constructed by this method are
                     full-length enriched cDNA library."
BASE COUNT       80 a     98 c     86 g     75 t
ORIGIN

  Query Match             87.6%;  Score 18.4;  DB 12;  Length 339;
  Best Local Similarity   95.0%;  Pred. No. 4.1e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              |||||||| ||||||||||
Db        104 TCATCCTCATCCTTGTCCTC 123


RESULT 15
CB119380
LOCUS       CB119380                379 bp    mRNA    linear   EST 28-JAN-2003
DEFINITION  K-EST0166136 L8SCK0 Homo sapiens cDNA clone L8SCK0-9-B05 5', mRNA
            sequence.
ACCESSION   CB119380
VERSION     CB119380.1  GI:27945184
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 9  row: B  column: 05
            High quality sequence stop: 379.
FEATURES             Location/Qualifiers
     source          1. .379
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="L8SCK0-9-B05"
                     /sex="M"
                     /cell_line="SCK"
                     /lab_host="Top10F'"
                     /clone_lib="L8SCK0"
                     /note="Organ: Liver; Vector: pT7T3-Pac; Site_1: EcoRI;
                     Site_2: NotI; The library was contributed by the Soares
                     laboratory and it was constructed as described by Bonaldo,
                     M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
                     6(9): 791-806. RNA was prepared from harvested cell
                     culture."
BASE COUNT       94 a     98 c     93 g     94 t
ORIGIN

  Query Match             87.6%;  Score 18.4;  DB 14;  Length 379;
  Best Local Similarity   95.0%;  Pred. No. 4.2e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              ||||||||| ||||||||||
Db         77 TCATCCTCATCCTTGTCCTC 96


Search completed: February 19, 2004, 23:14:32
Job time : 1575 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 21:14:23 ; Search time 46 Seconds
(without alignments)
201.501 Million cell updates/sec

Title: US-10-085-108-21_COPY_175_195
Perfect score: 21
Sequence: 1 TCATCCTCACCCTTGTCCTCA 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	17.8	84.8	1958	4	US-09-702-327-3	Sequence 3, Appli
c	2	17	81.0	325	3	US-09-006-353A-16	Sequence 16, Appl
c	3	17	81.0	325	4	US-09-573-986-16	Sequence 16, Appl
c	4	17	81.0	1392	3	US-09-006-353A-1	Sequence 1, Appli
c	5	17	81.0	1392	4	US-09-573-986-1	Sequence 1, Appli
c	6	16.8	80.0	1512	3	US-08-955-918C-8	Sequence 8, Appli
c	7	16.8	80.0	1512	4	US-08-697-766A-8	Sequence 8, Appli
c	8	16.8	80.0	1801	3	US-08-955-918C-6	Sequence 6, Appli
c	9	16.8	80.0	1801	4	US-08-697-766A-6	Sequence 6, Appli
c	10	16.8	80.0	1804	1	US-08-631-200-1	Sequence 1, Appli
c	11	16.8	80.0	1804	1	US-08-829-553-1	Sequence 1, Appli
c	12	16.8	80.0	1804	2	US-08-922-267A-1	Sequence 1, Appli
c	13	16.8	80.0	1804	2	US-08-936-707A-1	Sequence 1, Appli
c	14	16.8	80.0	1804	2	US-08-936-706A-1	Sequence 1, Appli
c	15	16.8	80.0	1804	3	US-09-248-203-1	Sequence 1, Appli
c	16	16.8	80.0	1804	3	US-09-406-071-1	Sequence 1, Appli
c	17	16.8	80.0	2119	1	US-08-630-592-1	Sequence 1, Appli
c	18	16.8	80.0	2119	1	US-08-714-991-1	Sequence 1, Appli
c	19	16.8	80.0	2119	3	US-09-032-365A-1	Sequence 1, Appli
c	20	16.8	80.0	2434	1	US-08-630-592-3	Sequence 3, Appli
c	21	16.8	80.0	2434	1	US-08-714-991-3	Sequence 3, Appli
c	22	16.8	80.0	2434	3	US-09-032-365A-3	Sequence 3, Appli
c	23	16.8	80.0	2518	3	US-09-433-699-3	Sequence 3, Appli
c	24	16.8	80.0	6452	3	US-08-836-325-9	Sequence 9, Appli
c	25	16.2	77.1	272	4	US-09-313-294A-1834	Sequence 1834, Ap
	26	16.2	77.1	654	4	US-09-107-532A-479	Sequence 479, App
c	27	16.2	77.1	1052	1	US-08-466-603-1	Sequence 1, Appli
c	28	16.2	77.1	1052	1	US-08-314-503A-1	Sequence 1, Appli
c	29	16.2	77.1	1052	1	US-08-468-066-1	Sequence 1, Appli
c	30	16.2	77.1	1052	2	US-08-466-717-1	Sequence 1, Appli
c	31	16.2	77.1	1052	3	US-08-466-743-1	Sequence 1, Appli
c	32	16.2	77.1	1052	5	PCT-US95-12414-1	Sequence 1, Appli
c	33	16.2	77.1	1690	4	US-09-620-312D-69	Sequence 69, Appl
c	34	16.2	77.1	1923	4	US-09-620-312D-77	Sequence 77, Appl
c	35	16.2	77.1	2736	4	US-09-220-132-66	Sequence 66, Appl
c	36	16.2	77.1	3279	4	US-09-439-313-382	Sequence 382, App
c	37	16.2	77.1	3279	4	US-09-352-616A-382	Sequence 382, App
	38	16.2	77.1	6048	4	US-09-634-920-3	Sequence 3, Appli
	39	16.2	77.1	6048	4	US-09-514-907A-1	Sequence 1, Appli
c	40	16.2	77.1	536165	4	US-09-214-808-1	Sequence 1, Appli
c	41	16	76.2	2800	3	US-09-183-245-1	Sequence 1, Appli
c	42	16	76.2	2800	4	US-09-618-932-1	Sequence 1, Appli
c	43	16	76.2	3607	1	US-08-647-351B-1	Sequence 1, Appli
c	44	16	76.2	3607	4	US-09-326-074-1	Sequence 1, Appli
	45	15.8	75.2	20	3	US-09-433-699-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-09-702-327-3/c
; Sequence 3, Application US/09702327
; Patent No. 6426220
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Lex M. Cowsert
; TITLE OF INVENTION: ANTISENSE MODULATION OF CALRETICULIN EXPRESSION
; FILE REFERENCE: RTS-0097
; CURRENT APPLICATION NUMBER: US/09/702,327
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; LENGTH: 1958
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(1362)
US-09-702-327-3

Query Match 84.8%; Score 17.8; DB 4; Length 1958;
Best Local Similarity 90.5%; Pred. No. 32;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 TCATCCTCACCCTTGTCCTCA 21
              ||||||||| | |||||||||
Db       1292 TCATCCTCATCTTTGTCCTCA 1272
```

RESULT 2
US-09-006-353A-16/c
; Sequence 16, Application US/09006353A
; Patent No. 6261801
; GENERAL INFORMATION:
; APPLICANT: WEI, YING-FEI
; APPLICANT: YU, GUO-LIANG
; APPLICANT: GENTZ, REINER
; APPLICANT: RUBEN, STEVEN
; TITLE OF INVENTION: TUMOR NECROSIS FACTOR RECEPTOR 5
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/006,353A
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BROOKES, ANDERS A
;       REGISTRATION NUMBER:  36,373
;       REFERENCE/DOCKET NUMBER:  PF341
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (301) 309-8504
;       TELEFAX:  (301) 309-8512
;   INFORMATION FOR SEQ ID NO:  16:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  325 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-09-006-353A-16

  Query Match             81.0%;  Score 17;  DB 3;  Length 325;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTC 17
              |||||||||||||||||
Db        225 TCATCCTCACCCTTGTC 209


RESULT 3
US-09-573-986-16/c
; Sequence 16, Application US/09573986
; Patent No. 6455040
; GENERAL INFORMATION:
;  APPLICANT:  Wei, Ying-Fei
;  APPLICANT:  Ni, Jian
;  APPLICANT:  Gentz, Reiner
;  APPLICANT:  Ruben, Steven
;  TITLE OF INVENTION:  Tumor Necrosis Factor Receptor 5
;  FILE REFERENCE:  1488.1280004
;  CURRENT APPLICATION NUMBER: US/09/573,986
;  CURRENT FILING DATE:  2000-05-18
;  NUMBER OF SEQ ID NOS:  27
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 16
;   LENGTH: 325
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-573-986-16

  Query Match             81.0%;  Score 17;  DB 4;  Length 325;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTC 17
              |||||||||||||||||
Db        225 TCATCCTCACCCTTGTC 209


RESULT 4
US-09-006-353A-1/c
; Sequence 1, Application US/09006353A
; Patent No. 6261801
;  GENERAL INFORMATION:
;    APPLICANT:  WEI, YING-FEI
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  GENTZ, REINER
;    APPLICANT:  RUBEN, STEVEN
;    TITLE OF INVENTION:  TUMOR NECROSIS FACTOR RECEPTOR 5
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HUMAN GENOME SCIENCES, INC.
;      STREET:  9410 KEY WEST AVENUE
;      CITY:  ROCKVILLE
;      STATE:  MD
;      COUNTRY:  US
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/006,353A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROOKES, ANDERS A
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PF341
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1392 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..959
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  183..260
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  261..959
US-09-006-353A-1

  Query Match             81.0%;  Score 17;  DB 3;  Length 1392;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTC 17
              |||||||||||||||||
Db       1326 TCATCCTCACCCTTGTC 1310


RESULT 5
US-09-573-986-1/c
; Sequence 1, Application US/09573986
; Patent No. 6455040
; GENERAL INFORMATION:
;  APPLICANT:  Wei, Ying-Fei
;  APPLICANT:  Ni, Jian
;  APPLICANT:  Gentz, Reiner
;  APPLICANT:  Ruben, Steven
;  TITLE OF INVENTION:  Tumor Necrosis Factor Receptor 5
;  FILE REFERENCE:  1488.1280004
;  CURRENT APPLICATION NUMBER: US/09/573,986
;  CURRENT FILING DATE:  2000-05-18
;  NUMBER OF SEQ ID NOS:  27
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 1392
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (183)..(959)
US-09-573-986-1
```

```
  Query Match             81.0%;  Score 17;  DB 4;  Length 1392;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTC 17
              |||||||||||||||||
Db       1326 TCATCCTCACCCTTGTC 1310


RESULT 6
US-08-955-918C-8/c
; Sequence 8, Application US/08955918C
; Patent No. 6268130
;  GENERAL INFORMATION:
;    APPLICANT:  Kleyn, Patrick, and Moore, Karen
;    TITLE OF INVENTION:  RP Compositions and Therapeutic and
;    TITLE OF INVENTION:  Diagnostic Uses Therefor
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/955,918C
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/697,766
;      FILING DATE:  29-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  MNI-007CPDV2CPA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1512 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1515
US-08-955-918C-8

  Query Match             80.0%;  Score 16.8;  DB 3;  Length 1512;
  Best Local Similarity   90.0%;  Pred. No. 84;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              |||||||||  |||  ||||||
Db        611 TCATCCTCATCCTCGTCCTC 592


RESULT 7
US-08-697-766A-8/c
; Sequence 8, Application US/08697766A
; Patent No. 6399760
;  GENERAL INFORMATION:
;    APPLICANT:  Gimeno, Carlos; Kleyn, Patrick; and Moore, Karen J.
;    TITLE OF INVENTION:  RP Compositions and Therapeutic and
;    TITLE OF INVENTION:  Diagnostic Uses Therefor
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/697,766A
;      FILING DATE:  29-AUG-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silveri, Jean M.
;      REGISTRATION NUMBER:  39,030
;      REFERENCE/DOCKET NUMBER:  MNI-007
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1512 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1512
US-08-697-766A-8

  Query Match             80.0%;  Score 16.8;  DB 4;  Length 1512;
  Best Local Similarity   90.0%;  Pred. No. 84;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              |||||||||  |||  ||||||
Db        611 TCATCCTCATCCTCGTCCTC 592


RESULT 8
US-08-955-918C-6/c
; Sequence 6, Application US/08955918C
; Patent No. 6268130
;  GENERAL INFORMATION:
;    APPLICANT:  Kleyn, Patrick, and Moore, Karen
;    TITLE OF INVENTION:  RP Compositions and Therapeutic and
;    TITLE OF INVENTION:  Diagnostic Uses Therefor
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/955,918C
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/697,766
```

```
;       FILING DATE:  29-AUG-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Mandragouras, Amy E.
;       REGISTRATION NUMBER:  36,207
;       REFERENCE/DOCKET NUMBER:  MNI-007CPDV2CPA
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)227-7400
;       TELEFAX:  (617)227-5941
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1801 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  139..1653
US-08-955-918C-6

  Query Match             80.0%;  Score 16.8;  DB 3;  Length 1801;
  Best Local Similarity   90.0%;  Pred. No. 86;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              ||||||||| ||| ||||||
Db        749 TCATCCTCATCCTCGTCCTC 730


RESULT 9
US-08-697-766A-6/c
; Sequence 6, Application US/08697766A
; Patent No. 6399760
;  GENERAL INFORMATION:
;     APPLICANT:  Gimeno, Carlos; Kleyn, Patrick; and Moore, Karen J.
;     TITLE OF INVENTION:  RP Compositions and Therapeutic and
;     TITLE OF INVENTION:  Diagnostic Uses Therefor
;     NUMBER OF SEQUENCES:  13
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;       STREET:  28 State Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/697,766A
;       FILING DATE:  29-AUG-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Silveri, Jean M.
;       REGISTRATION NUMBER:  39,030
;       REFERENCE/DOCKET NUMBER:  MNI-007
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)227-7400
;       TELEFAX:  (617)227-5941
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1801 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  139..1653
US-08-697-766A-6

  Query Match             80.0%;  Score 16.8;  DB 4;  Length 1801;
  Best Local Similarity   90.0%;  Pred. No. 86;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              ||||||||| ||| ||||||
Db        749 TCATCCTCATCCTCGTCCTC 730


RESULT 10
US-08-631-200-1/c
; Sequence 1, Application US/08631200
; Patent No. 5646040
;  GENERAL INFORMATION:
;     APPLICANT:  Kleyn, Patrick W.
;     APPLICANT:  Moore, Karen J.
;     TITLE OF INVENTION:  COMPOSITIONS FOR THE TREATMENT AND
;     TITLE OF INVENTION:  DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
;     NUMBER OF SEQUENCES:  59
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/631,200
;       FILING DATE:  12-APR-1996
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Coruzzi, Laura A.
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  7853-057
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-9741/8864
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1804 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  139..1653
US-08-631-200-1

  Query Match             80.0%;  Score 16.8;  DB 1;  Length 1804;
  Best Local Similarity   90.0%;  Pred. No. 86;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              ||||||||| ||| ||||||
Db        749 TCATCCTCATCCTCGTCCTC 730


RESULT 11
US-08-829-553-1/c
; Sequence 1, Application US/08829553
; Patent No. 5817762
;  GENERAL INFORMATION:
;     APPLICANT:  Kleyn, Patrick W.
;     APPLICANT:  Moore, Karen J.
```

```
;    TITLE OF INVENTION:  COMPOSITIONS FOR THE TREATMENT AND
;    TITLE OF INVENTION:  DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/829,553
;      FILING DATE:  28-MAR-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/631,200
;      FILING DATE:  12-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7853-057
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  139..1653
US-08-829-553-1

  Query Match             80.0%;  Score 16.8;  DB 1;  Length 1804;
  Best Local Similarity   90.0%;  Pred. No. 86;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              ||||||||| ||| ||||||
Db        749 TCATCCTCATCCTCGTCCTC 730


RESULT 12
US-08-922-267A-1/c
; Sequence 1, Application US/08922267A
; Patent No. 5861239
;  GENERAL INFORMATION:
;    APPLICANT:  Kleyn, Patrick W.
;    APPLICANT:  Moore, Karen J.
;    TITLE OF INVENTION:  COMPOSITIONS FOR THE TREATMENT AND
;    TITLE OF INVENTION:  DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
;    NUMBER OF SEQUENCES:  82
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/922,267A
;      FILING DATE:  2-SEP-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/829,553
;      FILING DATE:  28-MAR-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/631,200
;      FILING DATE:  12-APR-1996
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7853-085
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  139..1653
US-08-922-267A-1

  Query Match             80.0%;  Score 16.8;  DB 2;  Length 1804;
  Best Local Similarity   90.0%;  Pred. No. 86;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              ||||||||| ||| ||||||
Db        749 TCATCCTCATCCTCGTCCTC 730


RESULT 13
US-08-936-707A-1/c
; Sequence 1, Application US/08936707A
; Patent No. 5871931
;  GENERAL INFORMATION:
;    APPLICANT:  Kleyn, Patrick W.
;    APPLICANT:  Moore, Karen J.
;    TITLE OF INVENTION:  COMPOSITIONS FOR THE TREATMENT AND
;    TITLE OF INVENTION:  DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/936,707A
;      FILING DATE:  24-SEP-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7853-100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
```

```
;       TELEFAX:  (212) 869-9741/8864
;       TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1804 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  139..1653
US-08-936-707A-1

  Query Match             80.0%;  Score 16.8;  DB 2;  Length 1804;
  Best Local Similarity   90.0%;  Pred. No. 86;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              |||||||||  |||  ||||||
Db        749 TCATCCTCATCCTCGTCCTC 730


RESULT 14
US-08-936-706A-1/c
; Sequence 1, Application US/08936706A
; Patent No. 5876919
;  GENERAL INFORMATION:
;    APPLICANT:  Kleyn, Patrick W.
;    APPLICANT:  Moore, Karen J.
;    TITLE OF INVENTION:  COMPOSITIONS FOR THE TREATMENT AND
;    TITLE OF INVENTION:  DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds LLP
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/936,706A
;       FILING DATE:  24-SEP-1997
;       CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;       NAME:  Coruzzi, Laura A.
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  7853-099
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-9741/8864
;       TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1804 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  139..1653
US-08-936-706A-1

  Query Match             80.0%;  Score 16.8;  DB 2;  Length 1804;
  Best Local Similarity   90.0%;  Pred. No. 86;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              |||||||||  |||  ||||||
Db        749 TCATCCTCATCCTCGTCCTC 730


RESULT 15
US-09-248-203-1/c
; Sequence 1, Application US/09248203
; Patent No. 6043346
;  GENERAL INFORMATION:
;    APPLICANT:  Kleyn, Patrick W.
;    APPLICANT:  Moore, Karen J.
;    TITLE OF INVENTION:  COMPOSITIONS FOR THE TREATMENT AND
;    TITLE OF INVENTION:  DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds LLP
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/248,203
;       FILING DATE:
;       CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/936,707
;       FILING DATE:  24-SEP-1997
;    ATTORNEY/AGENT INFORMATION:
;       NAME:  Coruzzi, Laura A.
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  7853-100
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-9741/8864
;       TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1804 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  139..1653
US-09-248-203-1

  Query Match             80.0%;  Score 16.8;  DB 3;  Length 1804;
  Best Local Similarity   90.0%;  Pred. No. 86;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              |||||||||  |||  ||||||
Db        749 TCATCCTCATCCTCGTCCTC 730


Search completed: February 19, 2004, 23:16:13
Job time : 48 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 22:22:13 ; Search time 194 Seconds
(without alignments)
398.744 Million cell updates/sec

Title: US-10-085-108-21_COPY_175_195
Perfect score: 21
Sequence: 1 TCATCCTCACCCTTGTCCTCA 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	21	100.0	424	9	US-09-864-761-21436	Sequence 21436, A
c	2	21	100.0	921	13	US-10-027-632-100997	Sequence 100997,
c	3	21	100.0	921	14	US-10-027-632-100997	Sequence 100997,
	4	21	100.0	1041	14	US-10-085-108-21	Sequence 21, Appl
	5	18.4	87.6	449	11	US-09-918-995-4056	Sequence 4056, Ap
c	6	18.4	87.6	772	15	US-10-106-698-1761	Sequence 1761, Ap
	7	18.4	87.6	852	10	US-09-764-864-388	Sequence 388, App
c	8	18.4	87.6	1840	13	US-10-027-632-99044	Sequence 99044, A
c	9	18.4	87.6	1840	14	US-10-027-632-99044	Sequence 99044, A
c	10	18.4	87.6	2522	12	US-10-104-047-1604	Sequence 1604, Ap
c	11	17.8	84.8	431	11	US-09-918-995-15180	Sequence 15180, A
c	12	17.8	84.8	635	13	US-10-027-632-77261	Sequence 77261, A
c	13	17.8	84.8	635	13	US-10-027-632-313997	Sequence 313997,
c	14	17.8	84.8	635	14	US-10-027-632-77261	Sequence 77261, A
c	15	17.8	84.8	635	14	US-10-027-632-313997	Sequence 313997,
c	16	17.8	84.8	657	13	US-10-027-632-35572	Sequence 35572, A
c	17	17.8	84.8	657	14	US-10-027-632-35572	Sequence 35572, A
c	18	17.8	84.8	1251	11	US-09-828-000-1	Sequence 1, Appli
c	19	17.8	84.8	1251	13	US-10-405-588-1	Sequence 1, Appli
c	20	17.8	84.8	1899	12	US-10-439-703-7	Sequence 7, Appli
c	21	17.8	84.8	1920	11	US-09-906-393A-35	Sequence 35, Appl
c	22	17.8	84.8	1924	13	US-10-240-965-265	Sequence 265, App
c	23	17.8	84.8	1940	13	US-10-367-093-13	Sequence 13, Appl
c	24	17.8	84.8	1958	10	US-09-791-406-3	Sequence 3, Appli
c	25	17.8	84.8	110096	10	US-09-880-107-1542	Sequence 1542, Ap
c	26	17.4	82.9	432	9	US-09-864-761-4009	Sequence 4009, Ap
c	27	17	81.0	325	9	US-09-826-212-16	Sequence 16, Appl
c	28	17	81.0	325	15	US-10-186-643-16	Sequence 16, Appl
	29	17	81.0	635	13	US-10-027-632-229451	Sequence 229451,
	30	17	81.0	635	14	US-10-027-632-229451	Sequence 229451,
c	31	17	81.0	829	13	US-10-027-632-169165	Sequence 169165,
c	32	17	81.0	829	14	US-10-027-632-169165	Sequence 169165,
	33	17	81.0	837	13	US-10-027-632-164807	Sequence 164807,
	34	17	81.0	837	13	US-10-027-632-164808	Sequence 164808,
	35	17	81.0	837	14	US-10-027-632-164807	Sequence 164807,
	36	17	81.0	837	14	US-10-027-632-164808	Sequence 164808,
c	37	17	81.0	1392	9	US-09-826-212-1	Sequence 1, Appli
c	38	17	81.0	1392	15	US-10-186-643-1	Sequence 1, Appli
	39	17	81.0	2203	12	US-10-104-047-148	Sequence 148, App
	40	16.8	80.0	160	9	US-09-864-761-28709	Sequence 28709, A
c	41	16.8	80.0	160	13	US-10-029-386-14292	Sequence 14292, A
	42	16.8	80.0	242	10	US-09-878-574-10238	Sequence 10238, A
c	43	16.8	80.0	296	10	US-09-960-352-11416	Sequence 11416, A
c	44	16.8	80.0	309	12	US-10-242-535A-57249	Sequence 57249, A
c	45	16.8	80.0	394	10	US-09-796-692-8033	Sequence 8033, Ap

ALIGNMENTS

RESULT 1
US-09-864-761-21436/c
; Sequence 21436, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30

```
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21436
;    LENGTH: 424
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    OTHER INFORMATION: MAP TO AL023279.1
;    OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.6
;    OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.2
;    OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
;    OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =   6
;    OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.4
;    OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.9
;    OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.7
;    OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.5
;    OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.3
;    OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.3
;    OTHER INFORMATION: NT HIT: gi11420096, EVALUE 5.00e-21
;    OTHER INFORMATION: SWISSPROT HIT: P43363, EVALUE 5.00e-14
;    OTHER INFORMATION: EST_HUMAN HIT: BE729944.1, EVALUE 5.00e-13
US-09-864-761-21436

  Query Match             100.0%;  Score 21;  DB 9;  Length 424;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              |||||||||||||||||||||
Db        411 TCATCCTCACCCTTGTCCTCA 391


RESULT 2
US-10-027-632-100997/c
; Sequence 100997, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 100997
;   LENGTH: 921
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-100997

  Query Match             100.0%;  Score 21;  DB 13;  Length 921;
  Best Local Similarity   100.0%;  Pred. No. 6.8;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              |||||||||||||||||||||
Db        692 TCATCCTCACCCTTGTCCTCA 672


RESULT 3
US-10-027-632-100997/c
; Sequence 100997, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 100997
;   LENGTH: 921
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-100997

  Query Match             100.0%;  Score 21;  DB 14;  Length 921;
  Best Local Similarity   100.0%;  Pred. No. 6.8;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              |||||||||||||||||||||
Db        692 TCATCCTCACCCTTGTCCTCA 672


RESULT 4
US-10-085-108-21
; Sequence 21, Application US/10085108
; Publication No. US20020176865A1
;   GENERAL INFORMATION:
;        APPLICANT: LUCAS, Sophie; BOON-FALLEUR, Thierry
;        TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES CODING
;                            FOR
;                            TUMOR REJECTION ANTIGEN PRECURSORS OF MEMBERS OF THE MAGE-C
;                            MAGE-B FAMILIES AND USES THEREOF
;        NUMBER OF SEQUENCES: 26
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Fulbright & Jaworski L.L.P.
;             STREET: 666 Fifth Avenue
;             CITY: New York City
;             STATE: New York
;             COUNTRY: USA
;             ZIP: 10103
;        COMPUTER READABLE FORM:
```

```
;              MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
;              COMPUTER: IBM PS/2
;              OPERATING SYSTEM: PC-DOS
;              SOFTWARE: Wordperfect
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/10/085,108
;              FILING DATE: 01-Mar-2002
;              CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 09/501,104
;              FILING DATE: 09-Feb-2000
;              APPLICATION NUMBER: 09/468,433
;              FILING DATE: December 17, 1999
;              APPLICATION NUMBER: 09/066,281
;              FILING DATE: April 24, 1998
;              APPLICATION NUMBER: 08/845,528
;              FILING DATE: April 25, 1997
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Mary Anne Schofield
;              REGISTRATION NUMBER: 36,669
;              REFERENCE/DOCKET NUMBER: LUD 5611.1 JEL/MAS
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (212) 318-3100
;              TELEFAX: (212) 318-3400
;    INFORMATION FOR SEQ ID NO: 21:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1041 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: SINGLE
;              TOPOLOGY: linear
;         SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-10-085-108-21

  Query Match             100.0%;  Score 21;  DB 14;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 6.7;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              |||||||||||||||||||||
Db        175 TCATCCTCACCCTTGTCCTCA 195


RESULT 5
US-09-918-995-4056
; Sequence 4056, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4056
;   LENGTH: 449
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-918-995-4056

  Query Match             87.6%;  Score 18.4;  DB 11;  Length 449;
  Best Local Similarity   95.0%;  Pred. No. 94;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              |||||||||  ||||||||||
Db         89 TCATCCTCATCCTTGTCCTC 108


RESULT 6
US-10-106-698-1761/c
; Sequence 1761, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
;  FILE REFERENCE: PA005P1
;  CURRENT APPLICATION NUMBER: US/10/106,698
;  CURRENT FILING DATE:  2002-03-27
;  PRIOR APPLICATION NUMBER: PCT/US00/26524
;  PRIOR FILING DATE: 2000-09-28
;  PRIOR APPLICATION NUMBER: US 60/157,137
;  PRIOR FILING DATE: 1999-09-29
;  PRIOR APPLICATION NUMBER: US 60/163,280
;  PRIOR FILING DATE: 1999-11-03
;  NUMBER OF SEQ ID NOS: 8564
;  SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1761
;   LENGTH: 772
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (214)..(214)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (766)..(766)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (772)..(772)
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1761

  Query Match             87.6%;  Score 18.4;  DB 15;  Length 772;
  Best Local Similarity   95.0%;  Pred. No. 91;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 CATCCTCACCCTTGTCCTCA 21
              ||||||||  ||||||||||
Db        491 CATCCTCATCCTTGTCCTCA 472


RESULT 7
US-09-764-864-388
; Sequence 388, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 388
;   LENGTH: 852
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (621)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (834)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (836)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-388

  Query Match             87.6%;  Score 18.4;  DB 10;  Length 852;
  Best Local Similarity   95.0%;  Pred. No. 90;
```

```
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTC 20
              ||||||||| ||||||||||
Db        135 TCATCCTCATCCTTGTCCTC 154


RESULT 8
US-10-027-632-99044/c
; Sequence 99044, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 99044
;   LENGTH: 1840
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-99044

  Query Match             87.6%;  Score 18.4;  DB 13;  Length 1840;
  Best Local Similarity   95.0%;  Pred. No. 87;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 CATCCTCACCCTTGTCCTCA 21
              || ||||||||||||||||||
Db        905 CAGCCTCACCCTTGTCCTCA 886


RESULT 9
US-10-027-632-99044/c
; Sequence 99044, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 99044
;   LENGTH: 1840
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-99044

  Query Match             87.6%;  Score 18.4;  DB 14;  Length 1840;
  Best Local Similarity   95.0%;  Pred. No. 87;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 CATCCTCACCCTTGTCCTCA 21
              || ||||||||||||||||||
Db        905 CAGCCTCACCCTTGTCCTCA 886


RESULT 10
US-10-104-047-1604/c
; Sequence 1604, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20030236392A1el full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1604
;   LENGTH: 2522
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-104-047-1604

  Query Match             87.6%;  Score 18.4;  DB 12;  Length 2522;
  Best Local Similarity   95.0%;  Pred. No. 85;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 CATCCTCACCCTTGTCCTCA 21
              |||||||| ||||||||||
Db       2273 CATCCTCATCCTTGTCCTCA 2254


RESULT 11
US-09-918-995-15180/c
; Sequence 15180, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 15180
;   LENGTH: 431
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(431)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-15180

  Query Match             84.8%;  Score 17.8;  DB 11;  Length 431;
  Best Local Similarity   90.5%;  Pred. No. 1.7e+02;
```

```
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              ||||||||| | |||||||||
Db        201 TCATCCTCATCTTTGTCCTCA 181


RESULT 12
US-10-027-632-77261/c
; Sequence 77261, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 77261
;   LENGTH: 635
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-77261

  Query Match             84.8%;  Score 17.8;  DB 13;  Length 635;
  Best Local Similarity   90.5%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              |||||||||||||| | |||
Db        262 TCATCCTCACCCTTGGCTTCA 242


RESULT 13
US-10-027-632-313997/c
; Sequence 313997, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 313997
;   LENGTH: 635
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-313997

  Query Match             84.8%;  Score 17.8;  DB 13;  Length 635;
  Best Local Similarity   90.5%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              |||||||||||||| | |||
Db        262 TCATCCTCACCCTTGGCTTCA 242


RESULT 14
US-10-027-632-77261/c
; Sequence 77261, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 77261
;   LENGTH: 635
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-77261

  Query Match             84.8%;  Score 17.8;  DB 14;  Length 635;
  Best Local Similarity   90.5%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCTCA 21
              |||||||||||||| | |||
Db        262 TCATCCTCACCCTTGGCTTCA 242


RESULT 15
US-10-027-632-313997/c
; Sequence 313997, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
```

; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 313997
; LENGTH: 635
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-313997

Query Match 84.8%; Score 17.8; DB 14; Length 635;
Best Local Similarity 90.5%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 TCATCCTCACCCTTGTCCTCA 21
              ||||||||||||||| | |||
Db        262 TCATCCTCACCCTTGGCTTCA 242
```

Search completed: February 20, 2004, 00:50:04
Job time : 196 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 23:14:43 ; Search time 171 Seconds
(without alignments)
331.510 Million cell updates/sec

Title: US-10-085-108-21_COPY_175_195
Perfect score: 21
Sequence: 1 TCATCCTCACCCTTGTCCTCA 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 1588498

Minimum DB seq length: 0
Maximum DB seq length: 21

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21	100.0	21	25	ABX95021	Human MAGE-C3 expr
	2	15.8	75.2	20	22	AAC92581	Human nucleolin ph
c	3	14.4	68.6	20	22	AAI66887	Nucleotide sequenc
c	4	14.4	68.6	20	22	AAD19679	Heat shock protein
	5	14.2	67.6	20	24	AAD24940	Antisense primer,
c	6	14	66.7	20	21	AAA07645	HERG gene exon 11/
	7	13.4	63.8	19	21	AAA79801	Hepatitis B virus
	8	13.4	63.8	20	21	AAA79799	Hepatitis B virus
	9	13.4	63.8	20	21	AAA79800	Hepatitis B virus
	10	13.4	63.8	21	21	AAZ76077	Human biallelic ma
	11	13.2	62.9	20	22	AAF55056	PCR primer used to
c	12	13.2	62.9	20	25	ABT19438	Aspergillus fumiga
c	13	12.8	61.0	17	24	ABV90153	Human POSHL1 scann
c	14	12.8	61.0	17	24	ABV90154	Human POSHL1 scann
c	15	12.8	61.0	20	21	AAA79747	Hepatitis B virus
c	16	12.8	61.0	20	21	AAA79748	Hepatitis B virus
	17	12.8	61.0	20	22	AAC81362	Human Y-box bindin
	18	12.8	61.0	20	24	AAD39529	Human calreticulin
	19	12.8	61.0	21	19	AAX09234	Human biallelic po
c	20	12.8	61.0	21	19	AAV35465	Primer for human h
c	21	12.8	61.0	21	21	AAZ75280	Human biallelic ma
	22	12.6	60.0	20	20	AAX90361	Human p53 gene rev
	23	12.6	60.0	20	20	AAX90375	Human p53 gene rev
	24	12.6	60.0	20	20	AAX90389	Human p53 gene rev
	25	12.6	60.0	20	22	AAC92590	Human nucleolin ph
	26	12.6	60.0	20	24	ABA81915	Rat G-protein sero
	27	12.6	60.0	20	25	ABZ77254	Antisense oligonuc
c	28	12.6	60.0	21	9	AAN82113	Probe for DNA enco
c	29	12.6	60.0	21	19	AAV57673	Exon 6 of an ENaC
	30	12.6	60.0	21	24	ABX97284	Human NOV-associat
c	31	12.4	59.0	17	24	ABV90151	Human POSHL1 scann
c	32	12.4	59.0	17	24	ABV90152	Human POSHL1 scann
	33	12.4	59.0	18	22	AAH47580	Human Her-3 mRNA i
c	34	12.4	59.0	19	16	AAT00705	Human trkC recepto
	35	12.4	59.0	19	21	AAA79802	Hepatitis B virus
c	36	12.4	59.0	19	21	AAZ88855	Human trkC recepto
c	37	12.4	59.0	20	22	AAC91653	Human angiotensino
c	38	12.4	59.0	20	22	AAC91657	Human angiotensino
	39	12.4	59.0	20	25	ABZ23815	EGFR mRNA inhibiti
c	40	12.4	59.0	21	17	AAT31713	Campylobacter fetu
	41	12.4	59.0	21	19	AAV57677	Exon 8 of an ENaC
c	42	12.2	58.1	17	24	ABN08085	Human GDMLP-1 17-m
c	43	12.2	58.1	17	24	ABN08086	Human GDMLP-1 17-m
c	44	12.2	58.1	17	24	ABN08406	Human GDMLP-1 17-m
c	45	12.2	58.1	17	24	ABN08407	Human GDMLP-1 17-m

ALIGNMENTS

RESULT 1
ABX95021
ID ABX95021 standard; DNA; 21 BP.
XX
AC ABX95021;
XX
DT 05-JUN-2003 (first entry)
XX
DE Human MAGE-C3 expression pattern anlaysis RT-PCR sense primer.
XX
KW TRAP; ss; tumour rejection antigen precursor; cytolytic T-cell; CTL;
KW tumour; seminoma; bladder transitional-cell carcinoma; NSCLC; adaptor;
KW head-and-neck squamous-cell carcinoma; breast carcinoma; sarcoma;
KW cutaneous melanoma; nonsmall cell lung cancer; RT-PCR; primer; MAGE-C3;
KW human; reverse transcription.
XX
OS Homo sapiens.
XX
PN US2002176865-A1.
XX
PD 28-NOV-2002.
XX
PF 01-MAR-2002; 2002US-0085108.
XX
PR 09-FEB-2000; 2000US-0501104.
PR 25-APR-1997; 97US-0845528.
PR 24-APR-1998; 98US-0066281.
PR 17-DEC-1999; 99US-0468433.
XX
PA (LUCA/) LUCAS S.
PA (BOON/) BOON-FALLEUR T.

XX
PI Lucas S, Boon-Falleur T;
XX
DR WPI; 2003-328468/31.
XX
PT Novel isolated nucleic acid encoding tumor rejection antigen precursor
PT MAGE-C3, MAGE-B5, or MAGE-B6, useful as diagnostic probes to determine
PT presence of abnormal e.g., tumor cells expressing MAGE-C1, MAGE-B5 or
PT MAGE-B6 -
XX
PS Example 12; Page 13; 59pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule which encodes
CC a tumour rejection antigen precursor (TRAP) having an amino acid sequence
CC of a TRAP encoded by a fully defined MAGE-C3, MAGE-B5, or MAGE-B6
CC polynucleotide sequence. Also disclosed is a method which is useful for
CC determining presence of cytolytic T-cells specific for complexes of human
CC leukocyte antigen (HLA) and a peptide derived from the nucleic acid in a
CC cytotoxic T-lymphocyte (CTL)-containing sample. The nucleic acid is
CC useful as a diagnostic probe to determine the presence of abnormal
CC (tumour) cells such as seminoma, bladder transitional-cell carcinoma,
CC head-and-neck squamous-cell carcinoma, breast carcinoma, sarcoma,
CC cutaneous melanoma or nonsmall cell lung cancer (NSCLC) which express
CC MAGE-C1, MAGE-B5 or MAGE-B6. The nucleic acid is useful for diagnosing a
CC disorder characterised by expression of MAGE-C1, MAGE-B5 or MAGE-B6 TRAPs
CC or tumour rejection antigens (TRAs). The present sequence represents the
CC human MAGE-C3 expression pattern anlaysis reverse transcription (RT)-PCR
CC sense primer.
XX
SQ Sequence 21 BP; 3 A; 10 C; 1 G; 7 T; 0 other;

Query Match 100.0%; Score 21; DB 25; Length 21;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGTCCTCA 21
||||||||||||||||||||||
Db 1 TCATCCTCACCCTTGTCCTCA 21

RESULT 2
AAC92581
ID AAC92581 standard; DNA; 20 BP.
XX
AC AAC92581;
XX
DT 27-MAR-2001 (first entry)
XX
DE Human nucleolin phosphorothioate antisense oligonucleotide, SEQ ID NO:31.
XX
KW Human nucleolin; P92; C23; phosphoprotein; ribosome biogenesis;
KW ribosome transport; cytokinesis; nucleogenesis; cell proliferation;
KW cell growth; transcriptional repression; replication;
KW signal transduction; chromatin decondensation; Ag-NOR family;
KW nucleolin antibody; systemic connective tissue disease; SLE;
KW systemic lupus erythematosus;
KW scleroderma-like chronic graft versus host disease;
KW expression inhibition; tumour formation; cancer; inflammation;
KW immune disorder; phosphorothioate; antisense oligonucleotide; ss.
XX
OS Homo sapiens.
XX
PN US6165786-A.
XX
PD 26-DEC-2000.
XX
PF 03-NOV-1999; 99US-0433699.
XX
PR 03-NOV-1999; 99US-0433699.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowsert LM;
XX
DR WPI; 2001-079848/09.
XX
PT Novel antisense compound targeted to human nucleolin which specifically
PT hybridizes with and inhibits the expression of human nucleolin, useful
PT for modulating the expression of nucleolin in cells -
XX
PS Example 15; Column 41-42; 41pp; English.
XX
CC Sequences AAC92560-C92639 represent antisense oligonucleotides
CC targetted to the human nucleolin gene, which inhibit its expression.
CC The antisense oligonucleotides were designed to target different
CC regions of the human nucleolin mRNA, and were analysed for their effect
CC on nucleolin mRNA levels by quantitative real-time PCR. Nucleolin (also
CC known as P92 or C23) is the most abundant nucleolar phosphoprotein in
CC actively growing cells. Nucleolin primarily participates in ribosome
CC biogenesis and transport of ribosomal components, being able to
CC transiently bind to pre-ribosomes in the nucleolus via a
CC ribonucleoprotein consensus sequence. However, it has also been shown to
CC be involved in cytokinesis, nucleogenesis, cell proliferation and
CC growth, transcriptional repression, replication, signal transduction,
CC and chromatin decondensation. Nucleolin is a member of the Ag-NOR
CC (active ribosomal gene located in the nucleolar organiser region) family
CC of proteins which are markers of active ribosomal genes, and whose
CC expression is associated with the prediction of tumour growth rate. The
CC presence of antibodies against nucleolin are associated with systemic
CC connective tissue diseases such as systemic lupus erythematosus (SLE)
CC and scleroderma-like chronic graft versus host disease. The
CC oligonucleotides of the invention are useful for diagnosis, prevention
CC and treatment of conditions associated with nucleolin expression, such as
CC tumour formation, immune disorders and inflammation.
XX
SQ Sequence 20 BP; 2 A; 9 C; 1 G; 8 T; 0 other;

Query Match 75.2%; Score 15.8; DB 22; Length 20;
Best Local Similarity 89.5%; Pred. No. 1.9e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGTCCT 19
||||||||| ||| |||||
Db 2 TCATCCTCATCCTCGTCCT 20

RESULT 3
AAI66887/c
ID AAI66887 standard; DNA; 20 BP.
XX
AC AAI66887;
XX
DT 07-JAN-2002 (first entry)
XX
DE Nucleotide sequence of primer seq Id No. 27.
XX
KW Nucleic acid amplification; hybridization assay; interaction assay;
KW expression cloning; PCR primer; ss.
XX
OS Synthetic.
XX
PN WO200171027-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-EP03311.
XX
PR 24-MAR-2000; 2000EP-0106450.
XX
PA (MICR-) MICROMET AG.
XX
PI Zohlnhoefer D, Klein C;
XX
DR WPI; 2001-611514/70.
XX
PT A method for the amplification of mRNA from a sample, e.g. for the

PT preparation of in vitro surrogates for pathologically modified cells or
PT tissues -
XX
PS Examples; Page 170; 170pp; English.
XX
CC The invention provides a novel method for the amplification of mRNA from
CC a sample. The method comprises (1) generating cDNA from polyadenylated
CC RNA employing at least 1 primer hybridizing to the polyadenylated RNA and
CC comprising a 5' poly(C) or a 5' poly(G) flank; (2) (either): (a)
CC (optionally): (i) (if present) removing non-hybridized, surplus primer(s)
CC and/or surplus dNTPs; (ii) 3' tailing of the generated cDNA with a poly
CC (G) tail when in step (1) primer(s) comprising a 5' poly(C) flank was/
CC were employed or a poly(C) tail when in step (1) primer(s) comprising a
CC 5' poly(G) flank was/were employed; or (b) (optionally) 3' tailing of the
CC generated cDNA with a poly(G) tail when in step (1) primer(s) comprising
CC a 5' poly(C) flank was/were employed or a poly(C) tail when in step (1)
CC primer(s) comprising a 5' poly(G) flank was/were employed using an RNA-
CC ligase, irrespective of the presence or absence of surplus primer(s) and
CC /or surplus dNTPs; and (3) amplifying the tailed cDNA with a primer
CC hybridizing to the tails generated in step (2ai) or (2b). The amplified
CC cDNA obtained may be used for in vitro and/or in vivo expression and
CC preparation of mRNA transcripts (which may then be used in hybridization
CC assays (comprising hybridization to oligonucleotide arrays, cDNA arrays
CC and/or PNA arrays) and/or interaction assays (comprising interactions
CC with carbohydrates, lectins, ribozymes, proteins, peptides, antibodies
CC and/or aptamers)) and for sequence specific PCR, cDNA cloning,
CC subtractive hybridization cloning and/or expression cloning. Sequences
CC AAI66876-B90 represent primers used in the method of the invention.
XX
SQ Sequence 20 BP; 6 A; 3 C; 8 G; 3 T; 0 other;

Query Match 68.6%; Score 14.4; DB 22; Length 20;
Best Local Similarity 93.8%; Pred. No. 7.1e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGT 16
||||||||| ||||||
Db 17 TCATCCTCAGCCTTGT 2

RESULT 4
AAD19679/c
ID AAD19679 standard; DNA; 20 BP.
XX
AC AAD19679;
XX
DT 18-DEC-2001 (first entry)
XX
DE Heat shock protein 70B PCR primer #1 related to the invention.
XX
KW Inhibitor; interferon-gamma; IFN-gamma signalling pathway; therapy;
KW restenosis; coronary artery; carotid artery; femoralis artery;
KW aorta-coronary vein bypass; arterial bypass; venous bypass;
KW ballon angioplasty; stent implantation; vasotropic; PCR primer; ss.
XX
OS Unidentified.
XX
PN WO200170953-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-EP03312.
XX
PR 24-MAR-2000; 2000EP-0106468.
XX
PA (MICR-) MICROMET AG.
XX
PI Zohlnhoefer D, Baeuerle P, Klein C, Neumann F;
XX
DR WPI; 2001-616404/71.
XX
PT Use of an inhibitor of the interferon-gamma signaling pathway for
PT preparation of a pharmaceutical composition useful in the treatment or
PT prevention of restenosis -
XX
PS Example 5; Page 55; 151pp; English.
XX
CC The present invention relates to the use of an inhibitor of the
CC interferon-gamma (IFN-gamma) signalling pathway for the preparation of
CC a pharmaceutical composition for the treatment or prevention of
CC restenosis. The inhibitor is useful for the preparation of a
CC pharmaceutical composition useful in treating or preventing restenosis
CC which include restenosis of coronary arteries, carotid arteries,
CC femoralis arteries, aorta-coronary vein bypass, arterial bypass,
CC and/or venous bypass; and restenotic modification, where prevention
CC of restenotic modification is done before, during and/or after ballon
CC angioplasty and/or stent implantation and the restenosis or restenotic
CC modification is in-stent restenosis; and for treating and preventing
CC restenosis in a subject preferably human. The present sequence is a
CC heat shock protein 70B PCR primer which is used in aberrant gene
CC expression in human restenotic tissue used in the exemplification
CC of the invention.
XX
SQ Sequence 20 BP; 6 A; 3 C; 8 G; 3 T; 0 other;

Query Match 68.6%; Score 14.4; DB 22; Length 20;
Best Local Similarity 93.8%; Pred. No. 7.1e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCATCCTCACCCTTGT 16
||||||||| ||||||
Db 17 TCATCCTCAGCCTTGT 2

RESULT 5
AAD24940
ID AAD24940 standard; DNA; 20 BP.
XX
AC AAD24940;
XX
DT 12-MAR-2002 (first entry)
XX
DE Antisense primer, to analyse human P28 alpha gene expression modulation.
XX
KW Human; growth inhibitory gene; retinoid; retinoic acid response element;
KW RARE site; therapy; promyelocytic leukaemia; cancer chemoprevention;
KW cytostatic; proteasome activator PA28 subunit alpha; PA28alpha gene;
KW PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200192578-A2.
XX
PD 06-DEC-2001.
XX
PF 25-MAY-2001; 2001WO-US17161.
XX
PR 26-MAY-2000; 2000US-207535P.
XX
PA (UNII) UNIV ILLINOIS FOUND.
XX
PI Roninson IB, Dokmanovic M, Chang B;
XX
DR WPI; 2002-075474/10.
XX
PT Expression construct encoding cellular genes, under control of a
PT promoter regulated by retinoids and cells comprising the construct for
PT identifying compounds that induce expression of the genes useful in
PT treating cancer -
XX
PS Example 1; Page 19; 64pp; English.
XX
CC The patent discloses growth inhibitory genes induced by retinoids. The
CC invention also relates to recombinant expression constructs that express
CC a reporter gene under the transcriptional control of a promoter for a
CC gene which is expressed by retinoid induction. The promoter does not

```
CC   contain a retinoic acid response elements (RARE) site. The invention
CC   further relates to reagents and methods for identifying compounds other
CC   than retinoids that modulate the expression of cellular genes. These
CC   compounds are useful for treating cancers such as promyelocytic leukaemia
CC   and cancer chemoprevention. The present DNA sequence is a PCR primer
CC   which is used for analysing human proteasome activator PA28 subunit
CC   alpha (PA28alpha) gene expression modulation by treatment with retinoic
CC   acid.
XX
SQ   Sequence 20 BP; 1 A; 10 C; 0 G; 9 T; 0 other;

  Query Match             67.6%;  Score 14.2;  DB 24;  Length 20;
  Best Local Similarity   84.2%;  Pred. No. 8.6e+03;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCT 19
              ||||||| ||||| | ||
Db          2 TCATCCTCCCCCTTCTTCT 20


RESULT 6
AAA07645/c
ID   AAA07645 standard; DNA; 20 BP.
XX
AC   AAA07645;
XX
DT   19-JUN-2000  (first entry)
XX
DE   HERG gene exon 11/intron 11 junction sequence.
XX
KW   HERG; mutation; long QT syndrome; LQT syndrome; gene therapy;
KW   human; ss.
XX
OS   Homo sapiens.
XX
PN   WO200006772-A1.
XX
PD   10-FEB-2000.
XX
PF   20-JUL-1999;   99WO-US16337.
XX
PR   27-JUL-1998;   98US-0122847.
PR   06-JAN-1999;   99US-0226012.
XX
PA   (UTAH ) UNIV UTAH RES FOUND.
XX
PI   Keating MT,  Splawski I;
XX
DR   WPI; 2000-195319/17.
XX
PT   New isolated mutant HERG nucleic acids, useful for developing products
PT   for the diagnosis, prevention and treatment of long QT syndrome  -
XX
PS   Example 8; Page 71; 163pp; English.
XX
CC   The invention relates to a HERG protein having a mutation compared to
CC   wild-type HERG, and is useful for developing products for the diagnosis,
CC   prevention and treatment of long QT (LQT) syndrome. The products and
CC   methods can be used for the diagnosis of subjects with LQT syndrome.
CC   They can also be used to screen for drugs for treating or preventing LQT
CC   syndrome. The HERG nucleic acids can also be used for gene therapy and
CC   HERG peptides can be used for peptide therapy. Sequences AAA07624-653
CC   represent intron/exon junction sequences of the HERG gene.
XX
SQ   Sequence 20 BP; 5 A; 3 C; 11 G; 1 T; 0 other;

  Query Match             66.7%;  Score 14;  DB 21;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 CCTCACCCTTGTCC 18
              ||||||||||||||
Db         16 CCTCACCCTTGTCC 3


RESULT 7
AAA79801
ID   AAA79801 standard; DNA; 19 BP.
XX
AC   AAA79801;
XX
DT   20-NOV-2000  (first entry)
XX
DE   Hepatitis B virus related oligonucleotide probe #64.
XX
KW   Hepatitis B virus; HBV; Hepatitis A virus; HAV; probe; detection;
KW   mutation; high-density gene chip; ss.
XX
OS   Hepatitis B virus.
XX
PN   CN1252452-A.
XX
PD   10-MAY-2000.
XX
PF   24-SEP-1999;   99CN-0114460.
XX
PR   24-SEP-1999;   99CN-0114460.
XX
PA   (UYDO-) UNIV DONGNAN.
XX
PI   Sun X,  Lu Z,  Wang Y;
XX
DR   WPI; 2000-443233/39.
XX
PT   High-density gene chip making process -
XX
PS   Example 1; Fig 15; 19pp; Chinese.
XX
CC   The present invention describes a method which comprises making a high-
CC   density gene chip, specifically for making high-density micro-array of
CC   oligonucleotide probes. An oligonucleotide probe selecting process to
CC   seek preferentially length variable and coverage variable probes is
CC   provided to ensure identical cross melting temperature of probes to the
CC   maximum limit, and this can make the cross control of gene chip
CC   relatively simple and raise the reliability of the gene chip detecting
CC   results. The process proposes a specific probe selection method for
CC   detecting target sequence directly, detecting mutation in both specific
CC   and non-specific sites and a probe overall arrangement scheme. AAA79738
CC   to AAA80201 represent oligonucleotide probe sequences which are used in
CC   examples from the present invention.
XX
SQ   Sequence 19 BP; 5 A; 8 C; 2 G; 4 T; 0 other;

  Query Match             63.8%;  Score 13.4;  DB 21;  Length 19;
  Best Local Similarity   93.3%;  Pred. No. 1.8e+04;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 CATCCTCACCCTTGT 16
              ||||||||||| ||
Db          1 CATCCTCACCCTAGT 15


RESULT 8
AAA79799
ID   AAA79799 standard; DNA; 20 BP.
XX
AC   AAA79799;
XX
DT   20-NOV-2000  (first entry)
XX
DE   Hepatitis B virus related oligonucleotide probe #62.
XX
KW   Hepatitis B virus; HBV; Hepatitis A virus; HAV; probe; detection;
KW   mutation; high-density gene chip; ss.
XX
OS   Hepatitis B virus.
```

XX
PN CN1252452-A.
XX
PD 10-MAY-2000.
XX
PF 24-SEP-1999; 99CN-0114460.
XX
PR 24-SEP-1999; 99CN-0114460.
XX
PA (UYDO-) UNIV DONGNAN.
XX
PI Sun X, Lu Z, Wang Y;
XX
DR WPI; 2000-443233/39.
XX
PT High-density gene chip making process -
XX
PS Example 1; Fig 15; 19pp; Chinese.
XX
CC The present invention describes a method which comprises making a high-
CC density gene chip, specifically for making high-density micro-array of
CC oligonucleotide probes. An oligonucleotide probe selecting process to
CC seek preferentially length variable and coverage variable probes is
CC provided to ensure identical cross melting temperature of probes to the
CC maximum limit, and this can make the cross control of gene chip
CC relatively simple and raise the reliability of the gene chip detecting
CC results. The process proposes a specific probe selection method for
CC detecting target sequence directly, detecting mutation in both specific
CC and non-specific sites and a probe overall arrangement scheme. AAA79738
CC to AAA80201 represent oligonucleotide probe sequences which are used in
CC examples from the present invention.
XX
SQ Sequence 20 BP; 5 A; 8 C; 2 G; 5 T; 0 other;

Query Match 63.8%; Score 13.4; DB 21; Length 20;
Best Local Similarity 93.3%; Pred. No. 1.8e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CATCCTCACCCTTGT 16
||||||||||||| ||
Db 4 CATCCTCACCCTAGT 18

RESULT 9
AAA79800
ID AAA79800 standard; DNA; 20 BP.
XX
AC AAA79800;
XX
DT 20-NOV-2000 (first entry)
XX
DE Hepatitis B virus related oligonucleotide probe #63.
XX
KW Hepatitis B virus; HBV; Hepatitis A virus; HAV; probe; detection;
KW mutation; high-density gene chip; ss.
XX
OS Hepatitis B virus.
XX
PN CN1252452-A.
XX
PD 10-MAY-2000.
XX
PF 24-SEP-1999; 99CN-0114460.
XX
PR 24-SEP-1999; 99CN-0114460.
XX
PA (UYDO-) UNIV DONGNAN.
XX
PI Sun X, Lu Z, Wang Y;
XX
DR WPI; 2000-443233/39.
XX
PT High-density gene chip making process -
XX
PS Example 1; Fig 15; 19pp; Chinese.
XX
CC The present invention describes a method which comprises making a high-
CC density gene chip, specifically for making high-density micro-array of
CC oligonucleotide probes. An oligonucleotide probe selecting process to
CC seek preferentially length variable and coverage variable probes is
CC provided to ensure identical cross melting temperature of probes to the
CC maximum limit, and this can make the cross control of gene chip
CC relatively simple and raise the reliability of the gene chip detecting
CC results. The process proposes a specific probe selection method for
CC detecting target sequence directly, detecting mutation in both specific
CC and non-specific sites and a probe overall arrangement scheme. AAA79738
CC to AAA80201 represent oligonucleotide probe sequences which are used in
CC examples from the present invention.
XX
SQ Sequence 20 BP; 5 A; 8 C; 2 G; 5 T; 0 other;

Query Match 63.8%; Score 13.4; DB 21; Length 20;
Best Local Similarity 93.3%; Pred. No. 1.8e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CATCCTCACCCTTGT 16
||||||||||||| ||
Db 3 CATCCTCACCCTAGT 17

RESULT 10
AAZ76077
ID AAZ76077 standard; DNA; 21 BP.
XX
AC AAZ76077;
XX
DT 10-SEP-2001 (first entry)
XX
DE Human biallelic marker downstream amplification primer SEQ ID NO:10433.
XX
KW Human genome; biallelic marker; high density disequilibrium map;
KW genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW haplotyping; hybridisation; identification; characterisation;
KW amplification; single nucleotide polymorphism; SNP; PCR primer;
KW diagnosis; ss.
XX
OS Homo sapiens.
XX
PN WO9954500-A2.
XX
PD 28-OCT-1999.
XX
PF 21-APR-1999; 99WO-IB00822.
XX
PR 21-APR-1998; 98US-0082614.
PR 23-NOV-1998; 98US-0109732.
XX
PA (GEST) GENSET.
XX
PI Cohen D, Blumenfeld M, Chumakov I;
XX
DR WPI; 2000-013267/01.
XX
PT Novel biallelic markers used to construct a high density disequilibrium
PT map of the human genome -
XX
PS Claim 9; Page 2455; 2745pp; English.
XX
CC AAZ65654 to AAZ69578 represent human biallelic markers from the present
CC invention, which contain a polymorphic base at position 24 of their
CC nucleotide sequences. AAZ69579 to AAZ77440 represent amplification
CC primers for the biallelic markers. The biallelic markers of the
CC invention have a variety of uses: they can be used for high density
CC mapping of the human genome, and in complex association studies and
CC haplotyping studies which are useful in determining the genetic basis
CC for disease states. Compositions and methods of the invention can also

CC be useful for the identification of the targets for the development of
CC pharmaceutical agents and diagnostic methods, as well as the
CC characterisation of the differential efficacious responses to and side
CC effects from pharmaceutical agents acting on a disease as well as other
CC treatment.
CC N.B. The SEQ ID NOs 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297
CC and 3367, are not actually given a sequence in the Sequence Listing
CC from the present invention.
XX
SQ Sequence 21 BP; 3 A; 8 C; 2 G; 8 T; 0 other;

Query Match 63.8%; Score 13.4; DB 21; Length 21;
Best Local Similarity 93.3%; Pred. No. 1.8e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 CTCACCCTTGTCCTC 20
||||||||||| |||
Db 7 CTCACCCTTGTTCTC 21

RESULT 11
AAF55056
ID AAF55056 standard; DNA; 20 BP.
XX
AC AAF55056;
XX
DT 15-MAY-2001 (first entry)
XX
DE PCR primer used to amplify a fragment of the mumps genome.
XX
KW Encapsidation protein; transcription protein; replication protein;
KW cell targeting; gene therapy; attenuated virus; vaccine; mumps;
KW PCR primer; ss.
XX
OS Mumps virus.
XX
PN WO200109309-A2.
XX
PD 08-FEB-2001.
XX
PF 02-AUG-2000; 2000WO-US21192.
XX
PR 02-AUG-1999; 99US-0146664.
PR 23-JUN-2000; 2000US-0213654.
XX
PA (AMHP) AMERICAN HOME PROD CORP.
XX
PI Clarke DK, Johnson EJ, Sidhu MS, Udem SA;
XX
DR WPI; 2001-123320/13.
XX
PT Producing a recombinant mumps virus (MUV), useful as a mumps vaccine,
PT by transfecting or transforming a host cell with a transcription vector
PT comprising a MUV genome or antigenome, and an expression vector
PT encoding trans-acting proteins -
XX
PS Example 1; Page 37; 133pp; English.
XX
CC PCR primers AAF55055-56 were used to amplify a fragment of the Mumps
CC virus genome. The amplified fragment was used in the course of the
CC invention. The specification describes a method for producing a
CC recombinant mumps virus. The method comprises transfecting or
CC transforming, in a rescue composition media, a host cell with a
CC transcription vector comprising a genome or antigenome of mumps virus,
CC and an expression vector encoding trans-acting proteins (NP, P and L)
CC necessary for encapsidation, transcription and replication. The method
CC is carried out under conditions sufficient to permit the co-expression
CC of the vectors and the production of the recombinant virus. The
CC recombinant virus has an ability to induce long-lasting immunity with a
CC single dose and a relatively low level of genome recombination. The
CC recombinantly produced Mumps viruses are useful in antibody generation,
CC diagnostic, prophylactic and therapeutic applications, cell targeting,
CC gene therapy, mutant virus preparation and immunogenic composition
CC preparation. The method may also produce an attenuated virus for use
CC as a vaccine for preventing or ameliorating mumps infection.
XX
SQ Sequence 20 BP; 1 A; 11 C; 2 G; 6 T; 0 other;

Query Match 62.9%; Score 13.2; DB 22; Length 20;
Best Local Similarity 83.3%; Pred. No. 2.2e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CATCCTCACCCTTGTCCT 19
| ||||||||| |||| |
Db 2 CCTCCTCACCCCTGTCTT 19

RESULT 12
ABT19438/c
ID ABT19438 standard; DNA; 20 BP.
XX
AC ABT19438;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene #1796.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response; ds.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US13142.
XX
PR 23-APR-2001; 2001US-285697P.
PR 27-APR-2001; 2001US-287066P.
PR 05-JUN-2001; 2001US-295890P.
PR 09-JUL-2001; 2001US-303899P.
PR 31-AUG-2001; 2001US-316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of
PT Aspergillus fumigatus, useful for treating or preventing infections by
PT A. fumigatus, or for treating a non-infectious disease in a subject
PT e.g. cancer -
XX
PS Disclosure; Page -; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding